Niveau(x) d'enseignement: 0 Numéro EDUQ : 8170861
Centre d'intérêt: 32

Surprenant, Claude-Guy.

Taux de scolarisation au Québec : analyse de régression et projection, 1961 à 1981 / Claude-Guy Surprenant ; [pour le] Ministère de l'éducation, Direction générale de la planification, Démographie scolaire . -- [Québec] : Le Ministère, 1973.
iv, 47 f. : graph. ; 28 cm. -- (Document de travail / Service de la démographie scolaire du Québec ; 9.09)

Titre de la couv.

Cote Dewey: 371.219'714/S961t Cote LC: LB/2846

Descripteur(s): *Scolarisation *Population étudiante *Taux de scolarisation *Prévision *Démographie *Analyse statistique

Identificateur(s) de sujet: *Analyse de régression *Droite de régression multiple

La prévision des taux de scolarisation est un des facteurs déterminants dans la réalisation d'un modèle de prévision des effectifs scolaires. Ce document présente les possibilités et les limites de l'une des méthodes de calcul de ces taux: la droite de régression multiple.

L'étude précise la nature de la régression linéaire simple et examine son application pratique dans le cas des taux de scolarisation. Elle entreprend la vérification statistique de cette méthode, expose ses limites et souligne les précautions à envisager dans son utilisation pour fins de prévision.

L'ajustement par la droite de régression est applicable au calcul des taux de scolarisation dans la plupart des cas. Rien n'indique toutefois qu'une tendance linéaire dans le passé le soit encore dans le présent et le demeure dans le futur. Une telle méthode ne peut en effet prévoir des changements brusques de comportement. Aussi devrait-elle être utilisée pour des prévisions à court terme (2 ou 3 ans). A mesure que de nouvelles données seraient disponibles, il faudrait refaire les calculs pour chaque âge et, si nécessaire, corriger la prévision.

Type de document: r
Auteur de l'analyse: ar

E D U Q 8170861

MINISTÈRE DE L' ÉDUCATION
Direction Générale de la Planification

Démographie scolaire
DOCUMENT DE TRAVAIL D. T. 9.09

TAUX DE SCOLARISATION AU QUÉBEC: ANALYSE DE RÉGRESSION ET PROJECTION, 1961 À 1981

AOÛT 1973 CLAUDE-GUY SURPRENANT

E D U Q 8170861

Avant-propos

Le groupe "Statistiques et analyse démographique" de la Direction générale de la planification compte parmi ses projets, celui de prévoir la population scolaire suivant la méthode des taux de scolarisation. Ce projet doit fournir des prévisions de la clientèle scolaire suivant le niveau d'enseignement, les secteurs général et professionnel, l'âge et le sexe pour l'ensemble du Québec et chacune des régions administratives scolaires.

La prévision des taux de scolarisation est un des facteurs déterminants dans la réalisation de ce modèle de prévision des effectifs scolaires. Plusieurs méthodes sont utilisées pour prévoir les taux de scolarisation. Ce document présente les possibilités et les limites de l'une d'elles: la droite de régression multiple.

Table des matières

	Pages
Avant-propos	i
Liste des graphiques, tableaux et figures	iii
Introduction	1
Chapitre I: Régression linéaire simple	
a) historique	3
b) la régression linéaire	3
1° position du problème	3
2° solution du problème	5
3° qualité de l'ajustement	10
a) test relatif à β	10
b) test de linéarité	13
1° des classes égales de y pour chaque x	13
2° des classes inégales de y pour chaque x	15
4° Prévision	17
Chapitre II: Application pratique	
a) calcul d'une droite de régression	18
b) test relatif à β	22
c) prévision	23
d) complément	24
Conclusion	26

Liste des graphiques, tableaux et figures

		Pages
Graphique 1	- Taux de scolarisation à 4 ans, Québec 1961 à 1981	27
Graphique 2	- Taux de scolarisation à 5 ans, Québec 1961 à 1981	28
Graphique 3	- Taux de scolarisation à 6 ans, Québec 1961 à 1981	29
Graphique 4	- Taux de scolarisation à 7 ans, Québec 1961 à 1981	30
Graphique 5	- Taux de scolarisation à 8 ans, Québec 1961 à 1981	31
Graphique 6	- Taux de scolarisation à 9 ans, Québec 1961 à 1981	32
Graphique 7	- Taux de scolarisation à 10 ans, Québec 1961 à 1981	33
Graphique 8	- Taux de scolarisation à 11 ans, Québec 1961 à 1981	34
Graphique 9	- Taux de scolarisation à 12 ans, Québec 1961 à 1981	35
Graphique 10	- Taux de scolarisation à 13 ans, Québec 1961 à 1981	36
Graphique 11	- Taux de scolarisation à 14 ans, Québec 1961 à 1981	37
Graphique 12	- Taux de scolarisation à 15 ans, Québec 1961 à 1981	38
Graphique 13	- Taux de scolarisation à 16 ans, Québec 1961 à 1981	39
Graphique 14	- Taux de scolarisation à 17 ans, Québec 1961 à 1981	40

Liste des graphiques, tableaux et figures (suite)

		Pages
Graphique 15	- Taux de scolarisation à 18 ans, Québec 1961 à 1981	41
Graphique 16	- Taux de scolarisation à 19 ans, Québec 1961 à 1981	42
Graphique 17	- Taux de scolarisation à 20 ans, Québec 1961 à 1981	43
Graphique 18	- Taux de scolarisation à 21 ans, Québec 1961 à 1981	44
Graphique 19	- Taux de scolarisation à 22 ans, Québec 1961 à 1981	45
Graphique 20	- Taux de scolarisation à 23 ans, Québec 1961 à 1981	46
Graphique 21	- Taux de scolarisation à 24 ans, Québec 1961 à 1981	47

* * *

Tableau 1	- Analyse de la variance	11
Tableau 2	- Taux de scolarisation à 4 ans, Québec 1961 à 1981	19
Tableau 3	- Taux de scolarisation à 4 ans, Québec 1971 à 1981	20
Tableau 4	- Analyse de la variance	
Tableau 5	- Test relatif à B	25

* * *

Figure 1		6

INTRODUCTION

Si depuis quelques années, les prévisions à long terme sont devenues presqu'indispensables, il n'en demeure pas moins, aux yeux des profanes, que ce sont là des efforts stériles puisqu'elles sont toujours inexactes. Mais une mauvaise prévision vaut mieux que rien du tout.

C'est pour répondre à leurs détracteurs que les chercheurs ont voulu rendre leurs méthodes de prévisions de plus en plus exactes et rigoureuses; ils se sont tournés vers la statistique: une science exacte et précise et assez souple dans la manipulation.

Mais l'utilisation de certaines méthodes statistiques est sous-jacente à certaines hypothèses dont le non-initié ne voit pas la nécessité, ni la subtilité. A partir d'une étude déjà terminée sur les taux de scolarisation par âge au Québec de 1961 à 1981[1], nous allons essayer de voir si la méthode de la droite d'ajustement ou droite de régression linéaire aurait pu conduire aux mêmes prévisions que la méthode utilisée dans la publication citée (taux estimés).

1- DUFOUR, D. et AMYOT, M.; Les taux de scolarisation au Québec 1961-1981, Document 9.08, Ministère de l'Education, Québec, juin 1972.

Dans un premier temps, nous préciserons la nature de la régression linéaire simple, nous verrons ensuite son application pratique dans le cas des taux de scolarisation, puis nous nous attacherons à la vérification statistique d'une telle méthode. Nous exposerons enfin ses limites et nous mentionnerons les précautions à envisager dans son utilisation pour fins de prévision. Ceci nous amènera à juger de la pertinence de cette méthode pour la prévision des taux de scolarisation au Québec.

Chapitre 1: Régression linéaire simple

a) historique:

Le concept de régression est apparu au XIXe siècle, et fut employé pour la première fois par le biologiste Galton. Celui-ci comparant la taille des parents à celle des enfants, constata que des pères très grands avaient des enfants plus petits et, que des pères très petits avaient des enfants moins petits. Ainsi la taille des enfants tendait à revenir à la taille moyenne lorsque la taille des parents s'en écartait. Il y avait alors régression.

C'est ainsi que ce terme est passé dans le vocabulaire de la statistique pour désigner, non un recul ou un retour en arrière, mais seulement une méthode d'investigation particulière[1].

b) la régression linéaire simple:

1°) position du problème

Nous allons restreindre notre étude à deux caractères quantitatifs ou à deux variables x et y, dont

---

1- Introduction à la statistique 3. Marcel Bertrand, P.U.M. 1971, p.207.

l'une est aléatoire et l'autre non-aléatoire. Si la variable y dépend de la variable x, alors x est la variable non-aléatoire ou variable indépendante, et y est la variable aléatoire ou variable dépendante.

La méthode de la droite de régression linéaire est l'ajustement d'une droite à une distribution statistique à deux caractères x et y, qui permet de résumer, d'interpréter et de prévoir les variations de y en fonction de x seulement.

i) l'équation de la droite de régression théorique

$$Y_i = a + \beta X_i + \mathcal{E}_i$$

ii) les hypothèses nécessaires à la réalisation de l'équation de la droite

a) $\mathcal{E}_i$ est une variable aléatoire non-contrôlée dont l'espérance mathématique est nulle:

b) $E(\mathcal{E}_i, \mathcal{E}_j) = 0$ c'est-à-dire que les variables aléatoires $\mathcal{E}_i$ ne sont pas corrélées entre elles.

c) $V(\mathcal{E}_i)$ = constante = σ^2
la variance des $\mathcal{E}_i$ est égale à une constance

d) les $\mathcal{E}_i$ sont indépendantes de X

d) les $\mathcal{E}_i$ seront supposées $N(0, \sigma)$

Nous reviendrons ultérieurement sur le pourquoi de ces hypothèses.

2° Solution du problème

La droite de régression n'est qu'un ajustement analytique, et l'équation de la droite de régression et par le fait même, les paramètres α et β peuvent être obtenus par la méthode des moindres carrés ou par la méthode du maximum de vraisemblance.

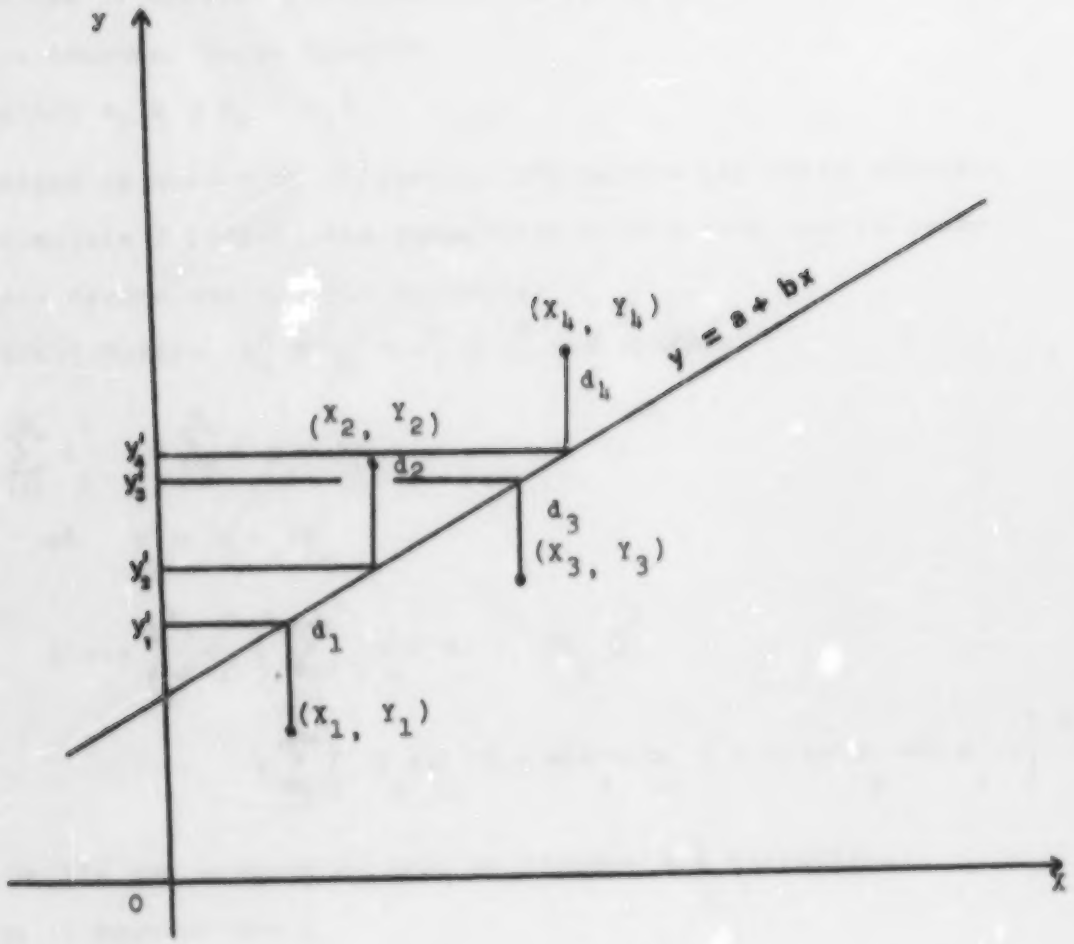
La méthode des moindres carrés que nous exposerons ici ne nécessite aucune hypothèse sur les $\mathcal{E}_i$, mais alors la prévision n'est qu'une estimation ponctuelle; alors que si on fait les hypothèses sur les $\mathcal{E}_i$, on peut estimer et tester les paramètres α , β et σ^2 .

Dans le cas expérimental l'équation de la droite de régression devient

$$y_i = a + bx_i$$

La méthode des moindres carrés consiste à minimiser la somme des carrés des écarts entre la droite de régression et les points expérimentaux (xi, yi). (Cf. figure 1).

figure 1



Menons un segment perpendiculaire à l'axe ox entre la droite de régression et les points; et soit d_i la longueur de ce segment.

alors $d_i = |Y_i - Y_i'|$

Ainsi la droite de régression déterminée par cette méthode, consiste à choisir des paramètres a et b tels que la somme des carrés des d_i soit minimale.

C'est-à-dire: $d_1^2 + d_2^2 + d_3^2 + d_4^2$ est minimal

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - y'_i)^2$$

où $y'_i = a + bx_i$

$$\text{alors } \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

$$= \sum_{i=1}^{n} \left[(y_i - m_y) - b(x_i - m_x) - (a - m_y + b m_x) \right]^2$$

On n'a que rajouté de part et d'autre des paramètres:

m_x : moyenne des x_i

m_y : moyenne des y_i

en effet : $\left[(y_i - m_y) - b(x_i - m_x) - (a - m_y + bm_x) \right]$

$$\left[y_i - m_y - bx_i + bm_x - a + m_y - bm_x \right]$$

$$\left[y_i - a - bx_i \right]$$

donc :

$$\sum_{i=1}^{n} \left[(y_i - m_y) - b(x_i - m_x) - (a - m_y + bm_x) \right]^2$$

$$= \sum_{i=1}^{n} \left[(y_i - m_y)^2 - 2b(x_i - m_x)(y_i - m_y) - 2(y_i - m_y)(a - m_y + bm_x) \right.$$
$$\left. + b^2 (x_i - m_x)^2 + 2b(x_i - m_x)(a - m_y + bm_x) + (a - m_y + bm_x)^2 \right]$$

$$= \sum_{i=1}^{n} (y_i - m_y)^2 - 2b \sum_{i=1}^{n} (x_i - m_x)(y_i - m_y) - 2(a - m_y + bm_x) \sum_{i=1}^{n} (y_i - m_y)$$
$$+ b^2 \sum_{i=1}^{n} (x_i - m_x)^2 + 2b(a - m_y + bm_x) \sum_{i=1}^{n} (x_i - m_x) + \sum_{i=1}^{n} (a - m_y + bm_x)^2$$

$$= \sum_{i=1}^{n} (y_i - m_y)^2 - 2b \sum_{i=1}^{n} (y_i - m_y)(x_i - m_x) + b^2 \sum_{i=1}^{n} (x_i - m_x)^2 + n(a - m_y + bm_x)^2$$

car : $\sum_{i=1}^{n} (x_i - m_x) = 0$ et $\sum_{i=1}^{n} (y_i - m_y) = 0$

$$= ns_y^2 - 2bns_{xy} + b^2 ns_x^2 + n(a - m_y + bm_x)^2 \qquad (-I-)$$

car: $$s_y^2 = \sum_{i=1}^{n} \frac{(y_i - m_y)^2}{n}$$

$$s_x^2 = \sum_{i=1}^{n} \frac{(x_i - m_x)^2}{n}$$

$$s_{xy} = \sum_{i=1}^{n} \frac{(x_i - m_x)(y_i - m_y)}{n}$$

On minimise les écarts, en égalant la dérivée première de l'équation (-I-), par rapport à a et à b, à zéro.

donc: $$\frac{\partial}{\partial a} = 2n(a - m_y + bm_x) = 0$$

$$= a - m_y + bm_x = 0$$

alors: $$a = m_y - bm_x$$

$$\boxed{a = m_y - bm_x}$$

de même : $\frac{\partial}{\partial b} = -2ns_{xy} + 2nbs_x^2 = 0$

$$2nbs_x^2 = 2ns_{xy}$$

alors : $$b = \frac{s_{xy}}{s_x^2}$$

$$\boxed{b = \frac{s_{xy}}{s_x^2}}$$

On a ainsi calculé les deux paramètres de la droite de régression qui donne le meilleur ajustement analytique.

3°) <u>qualité de l'ajustement</u>

a) test relatif à β

Il existe plusieurs tests statistiques pour vérifier la valeur significative du β .

Avec les hypothèses admises plus haut, il est possible de tester l'hypothèse selon laquelle $\beta = 0$ (absence de régression)

L'équation initiale d'analyse de la variance:

$$\sum_{i=1}^{n}\left[y_i - m_y\right]^2 = \sum_{i=1}^{n}\left[y_i - y'_i\right]^2 + \sum_{i=1}^{n}\left[y'_i - m_y\right]^2$$

s'interprète comme suit: le premier terme, ou somme des carrés des écarts à la moyenne, correspond à la variance marginale de y; le second terme, ou somme des carrés des distances d_i correspond à une variance résiduelle autour de la droite de régression; enfin, le dernier terme, ou somme des carrés des distances entre les points de la droite de régression et de la moyenne M_y, correspond à la variance expliquée par la droite de régression.

Le tableau d'analyse de la variance se présente ainsi

Tableau 1 - Analyse de la variance

VARIATION	SOMME DES CARRES	DEGRES DE LIBERTE	ESTIMATION DE σ^2
DUE A LA REGRESSION	$\sum_{i=1}^{n}\left[y'_i - m_y\right]^2$	1	$V_A = b^2 n s_x^2$
RESIDUELLE	$\sum_{i=1}^{n}\left[y_i - y'_i\right]^2$	n - 2	$V_r = \sum_{i=1}^{n}\left[y_i - y'_i\right]^2 / n-2$
TOTALE	$\sum_{i=1}^{n}\left[y_i - m_y\right]^2$	n - 1	$F'_{n-2} = \frac{V_A}{V_r}$

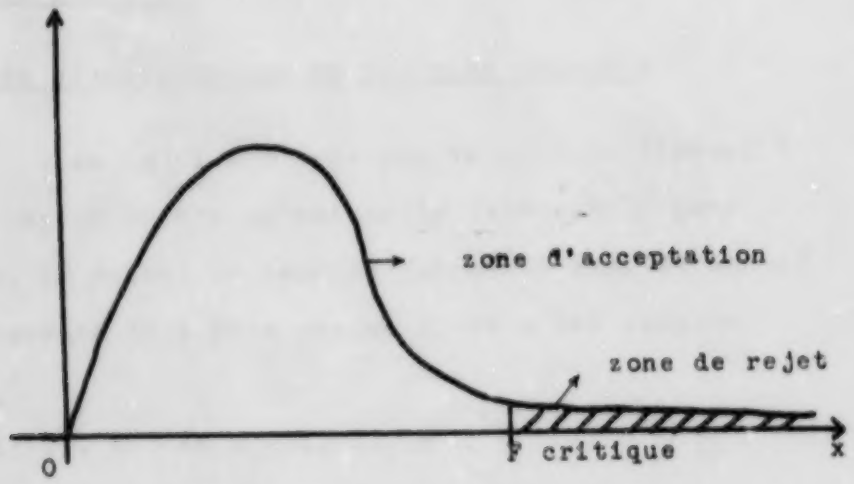
i) Ho: $\beta = 0$

ii) $\alpha = 0.05$

iii) F critique $= F'_{n-2,0.05}$

ıvi) F expérimental $= V_A / V_r$

vi) décision



Selon que le F expérimental se situe à gauche ou à droite du F critique on accepte ou on rejette l'hypothèse Ho. Si on acceptait l'hypothèse que $\beta = 0$, ceci impliquerait que le coefficient de corrélation linéaire de Pearson, serait égal lui aussi à zéro.

en effet $b = \dfrac{s_{xy}}{s_x^2}$

$$r = \frac{s_{xy}}{s_x \, s_y} = \frac{b \, s_x}{s_y}$$

Donc il n'y aurait pas de liaison linéaire entre la variable y et la variable x (à moins d'une distribution circulaire). Il faudrait trouver entre ces deux variables une relation exponentielle, logarithmique ou autre.

b) <u>test de linéarité</u>

1°) <u>des classes égales de y pour chaque x</u>

Bien qu'il n'existe pas de test de linéarité dans le cas où on n'a qu'une seule valeur de y pour chaque x, il existe un test de linéarité dans le cas où on a k valeurs de y pour chaque x. On a les couples (X_i, Y_{ih})

$i = 1, \ldots n$ et $h = 1, \ldots k$

l'équation d'analyse de la variance devient:

$$\sum_{i=1}^{n}\sum_{h=1}^{k}\left[y_{ih}-m_y\right]^2 = \sum_{i=1}^{n}\sum_{h=1}^{k}\left[y_{ih}-y_{i\bullet}\right]^2 + k\sum_{i=1}^{n}\left[y_{i\bullet}-y'_i\right]^2 + \sum_{i=1}^{n}k\left[y'_i-m_y\right]^2$$

en divisant chaque terme par σ^2 et par l'application du théorème de Cochran, on obtient

$$\frac{\sum_{i=1}^{n}\sum_{h=1}^{k}\left[y_{ih}-m_y\right]}{\sigma^2} = \frac{\sum_{i=1}^{n}\sum_{h=1}^{k}\left[y_{ih}-y_{i\bullet}\right]}{\sigma^2} + \frac{k\sum_{i=1}^{n}\left[y_{i\bullet}-y'_i\right]}{\sigma^2} + \frac{\sum_{i=1}^{n}k\left[y'_i-m_y\right]}{\sigma^2}$$

$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$

$$\chi^2_{nk-1} \qquad \chi^2_{nk-n} \qquad \chi^2_{n-2} \qquad \chi^2_1$$

i) Ho: linéarité

ii) $\alpha = 0.05$

iii) F critique $= F^{I}_{nk-n,0.05}$

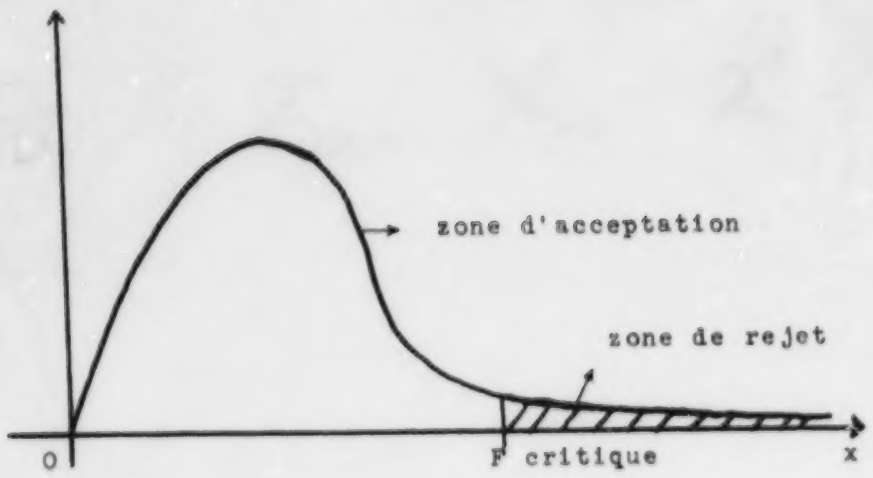
Ivi) F expérimental $= V_A / V_r$

$$V_A = \frac{\sum_{i=1}^{n}\sum_{h=1}^{k}(y_{ih}-y_{i.})^2}{nk-n}$$

$$V_r = \frac{k\sum_{i=1}^{n}(y_{i.}-y'_i)^2}{n-2}$$

$$y_{i.} = \frac{\sum_{h=1}^{k} y_{ih}}{k}$$

vi) décision



2° <u>des classes inégales de y pour chaque x</u>

Il existe aussi un test de linéarité lorsqu'on a des classes inégales de y pour chaque x. Nous avons les couples suivants: (X_i, Y_{ih_i})

$h_i = 1, 2, \ldots, k_i$

$i = 1, 2, \ldots n$

l'équation d'analyse de la variance est la suivante:

$$\sum_{i=1}^{n}\sum_{h_i=1}^{k_i}\left[y_{ih_i}-m_y\right]^2=\sum_{i=1}^{n}\sum_{h_i=1}^{k_i}\left[y_{ih_i}-y_{i\bullet}\right]^2+\sum_{i=1}^{n}\left[k_i\left(y_{i\bullet}-y'_i\right)\right]^2+\sum_{i=1}^{n}\left[k_i\left(y'_i-m_y\right)\right]^2$$

en divisant chaque membre par σ^2 et par l'application du théorème de Cochran.

$$\frac{\sum_{i=1}^{n}\sum_{h_i=1}^{k_i}\left[y_{ih_i}-m_y\right]^2}{\sigma^2}=\frac{\sum_{i=1}^{n}\sum_{h_i=1}^{k_i}\left[y_{ih_i}-y_{i\bullet}\right]^2}{\sigma^2}+\frac{\sum_{i=1}^{n}\left[k_i\left(y_{i\bullet}-y'_i\right)\right]^2}{\sigma^2}+\frac{\sum_{i=1}^{n}\left[k_i\left(y'_i-m_y\right)\right]^2}{\sigma^2}$$

$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$

$$\chi^2_{\sum k_i} \qquad \chi^2_{\sum k_i - n} \qquad \chi^2_{n-2} \qquad \chi^2_1$$

i) Ho: linéarité

ii) $a = 0.05$

iii) F critique $= F^{I}_{\sum k_j \,.\, 0.05}$

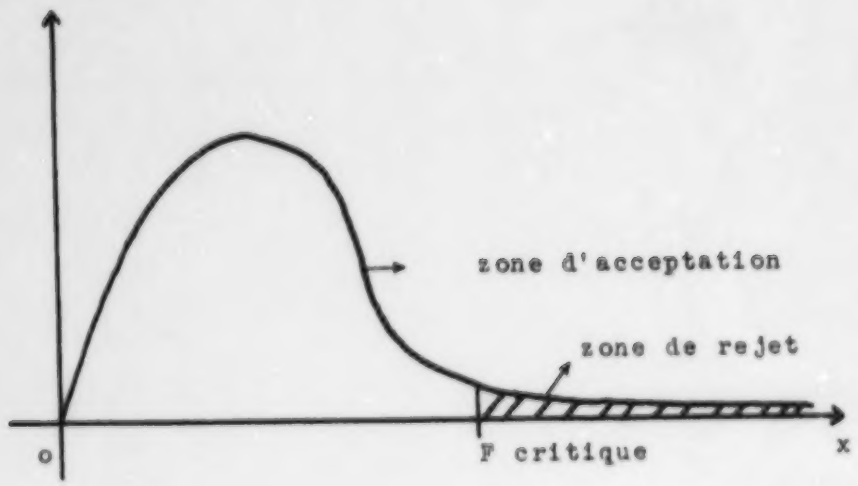
Ivi) F expérimental $= V_A \,/\, V_r$

$$V_A = \frac{\sum_{i=1}^{n}\sum_{h_i=1}^{k_i}(y_{ih_i} - y_{i.})^2}{\sum_{i=1}^{n} k_i - n}$$

$$V_r = \frac{\sum_{i=1}^{n} k_i (y_{i.} - y'_i)^2}{n-2}$$

$$y_{i.} = \frac{\sum_{h_i=1}^{k_i} y_{ih_i}}{k_i}$$

vi) décision



4°) Prévision

La droite de régression étant connue, il est possible de faire des estimations des valeurs que peut prendre y lorsque l'on connaît les valeurs de x.

C'est ainsi que l'ont peut prévoir la valeur moyenne de Y'_o pour un x_o donné, par la formule suivante:

$$y'_o = a + b x_o \pm k_a \sqrt{1 + \frac{(x_o - m_x)^2}{s_x^2}} \quad \frac{\sum_{i}^{n} d_i^2}{n(n-2)}$$

la valeur de k_a est tirée du $t_{n-2,a}$. On obtient une estimation par intervalle de la valeur moyenne de Y'_o avec un léger écart; c'est-à-dire $Y'_o = a + b x_o$ avec ± un certain écart dû à la régression.

Chapitre 2: Application pratique

a) Calcul d'une droite de régression

Comme nous l'avons déjà mentionné, nous disposons des taux de scolarisation par âge, observés pour les années 1961 à 1968 et estimés pour les années 1969, 1970 et 1971[1]. Considérons les taux de scolarisation à 4 ans dont les données figurent au tableau 2.

Peut-on appliquer la méthode de la droite de régression à cette série chronologique? A l'aide de certains artifices de calcul, nous calculerons la droite de régression et nous vérifierons sa linéarité par la suite.

Nous attribuons les valeurs arbitraires à x de telle sorte que la $\sum_{i=1}^{n} x_i = 0$ pour simplifier les calculs et les rendre moins lourds (Cf. tableau 3)

1- Ibid. page 16.

Tableau 2: Taux de scolarisation à 4 ans. Québec 1961 à 1971

ANNEE	1961	62	63	64	65	66	67	68	69	70	71
TAUX DE SCOLARISATION	0.038	0.039	0.051	0.092	0.128	0.166	0.184	0.210	0.230	0.248	0.249

Tableau 3: Taux de scolarisation à 4 ans, Québec 1961-71

année	X_i	Y_i taux de scolarisation	X_i^2	$X_i Y_i$
1961	-5	0.038	25	-0.190
1962	-4	0.039	16	-0.156
1963	-3	0.051	9	-0.153
1964	-2	0.092	4	-0.128
1965	-1	0.128	1	0.0
1966	0	0.166	0	0.184
1967	1	0.184	1	0.420
1968	2	0.210	4	0.690
1969	3	0.230	9	0.992
1970	4	0.248	16	1.245
1971	5	0.249	25	
TOTAL	0	1.635	110	2.72

Notons que n = 11 puisque nous avons 11 observations expérimentales.

$$a = m_y - b m_x = \frac{\sum_{i=1}^{n} y_i}{n} - b \frac{\sum_{i=1}^{n} x_i}{n}$$

$$b = \frac{s_{xy}}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i\, y_i - m_x\, m_y}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - m_x^2}$$

puisque $\sum_{i=1}^{n} x_i = 0$ donc $m_x = 0$

alors $a = \frac{\sum_{i=1}^{n} y_i}{n} = m_y$

$$b = \frac{\sum_{i=1}^{n} x_i\, y_i}{\sum_{i=1}^{n} x_i^2}$$

ainsi $a = 1.635 / 11 = 0.149$

$b = 2.72 / 110 = 0.025$

alors l'équation de la droite de régression devient:

$y = 0.149 + 0.025\ x$

b) <u>test relatif à β</u>

Tableau 4: Analyse de la variance

Somme des carrés	degré de liberté	Estimation de σ^2
$\sum_{i=1}^{n}\left[y'_i - m_y\right]^2 = 0.237$	1	$V_A = 0.2369$
$\sum_{i=1}^{n}\left[y_i - y'_i\right]^2 = 0.007$	9	$V_r = 0.0008$
$\sum_{i=1}^{n}\left[y_i - m_y\right]^2 = 0.244$	10	$F = \frac{V_A}{V_r} = 29.1$

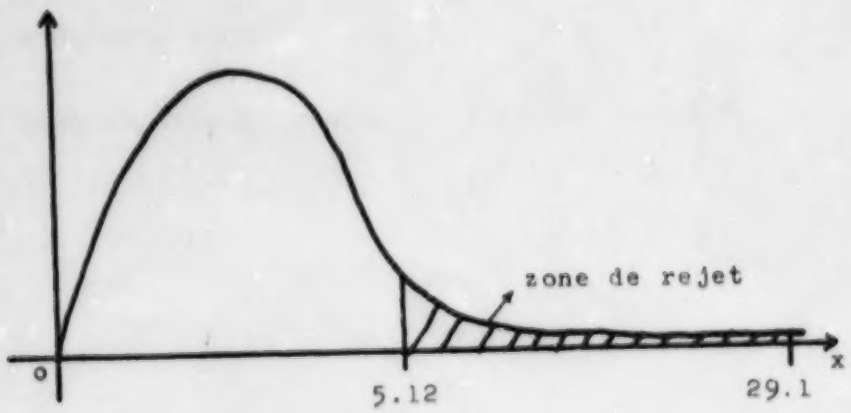
i) $H_0: \beta = 0$

ii) $\alpha = 0.05$

iii) $F^1_{9,0.05} = 5.12$

iv) F expérimental $= V_A / V_r = 29.1$

v) décision



On rejette l'hypothèse Ho à savoir que $\beta = 0$, β diffère significativement de zéro.

C- Prévision

Quelle sera la valeur moyenne de Y'_o en l'an 1976 ($X_o = 10$)?

$$y'_o = a + b\,x_o \pm k_a \sqrt{1 + \frac{(x_o - m_x)^2}{s_x^2}} \quad \frac{\sum_{i=1}^{n} d_i}{n(n-2)}$$

$$\sum_{i=1}^{11} d_1^2 = \sum_{i=1}^{11} (y_1 - y'_i)^2 = 0.007$$

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{11} x_1^2 - m_x^2 = \frac{1}{11}\sum_{i=1}^{11} x_1^2 = \frac{110}{11} = 10$$

$$Y'_o = 0.149 + 0.025\,(10) \pm 2.262\sqrt{1 + \frac{(10)^2}{10}\,\frac{0.007}{11(9)}}$$

$$Y'_o = 0.399 \pm 0.03$$

Y'_o est compris dans l'intervalle (0.369, 0.429).

d) complément

Or a repris chaque taux de scolarisation par âge pour la période 1961 à 1971, puis on a calculé la droite de régression, et fait pour chacun des taux le test relatif à β. Les résultats sont compilés au tableau 5.

On trouvera aussi en annexe des graphiques des taux de scolarisation (estimés) pour la période 1961 à 1981, ainsi que la droite de régression calculée sur la période de 1961 à 1971 et sa projection jusqu'en 1981, de même que l'équation de la droite de régression.

Tableau 5 Test relatif à B

Age	F critique	F expérimental	$\beta = 0$ (absence de régression)
4	5.12	29.10	NON
5		276.60	NON
6		3.90	OUI
7		2.92	
8		0.10	
9		1.68	
10		1.84	
11		0.97	
12		3.77	OUI
13		13.72	NON
14		39.26	
15		215.65	
16		440.19	
17		1,110.90	
18		1,118.30	
19		365.37	
20	5.12	33.24	
21	6.61	114.52	
22		109.18	
23		45.5	
24		47.0	NON

CONCLUSION

Bien que l'ajustement par la droite de régression soit applicable au calcul des taux dans la plupart des cas, (là où elle ne s'applique pas, la cause pourrait être une mauvaise déclaration d'âge [1], il faut être extrêmement prudent quant à son prolongement dans le futur.

Si les taux de scolarisation par âge sont liés à l'année en cours, ils le sont indirectement, puisque ce taux est en fait fonction du nombre d'étudiants fréquentant l'école à temps complet à un âge donné et du nombre total de personnes du même âge.

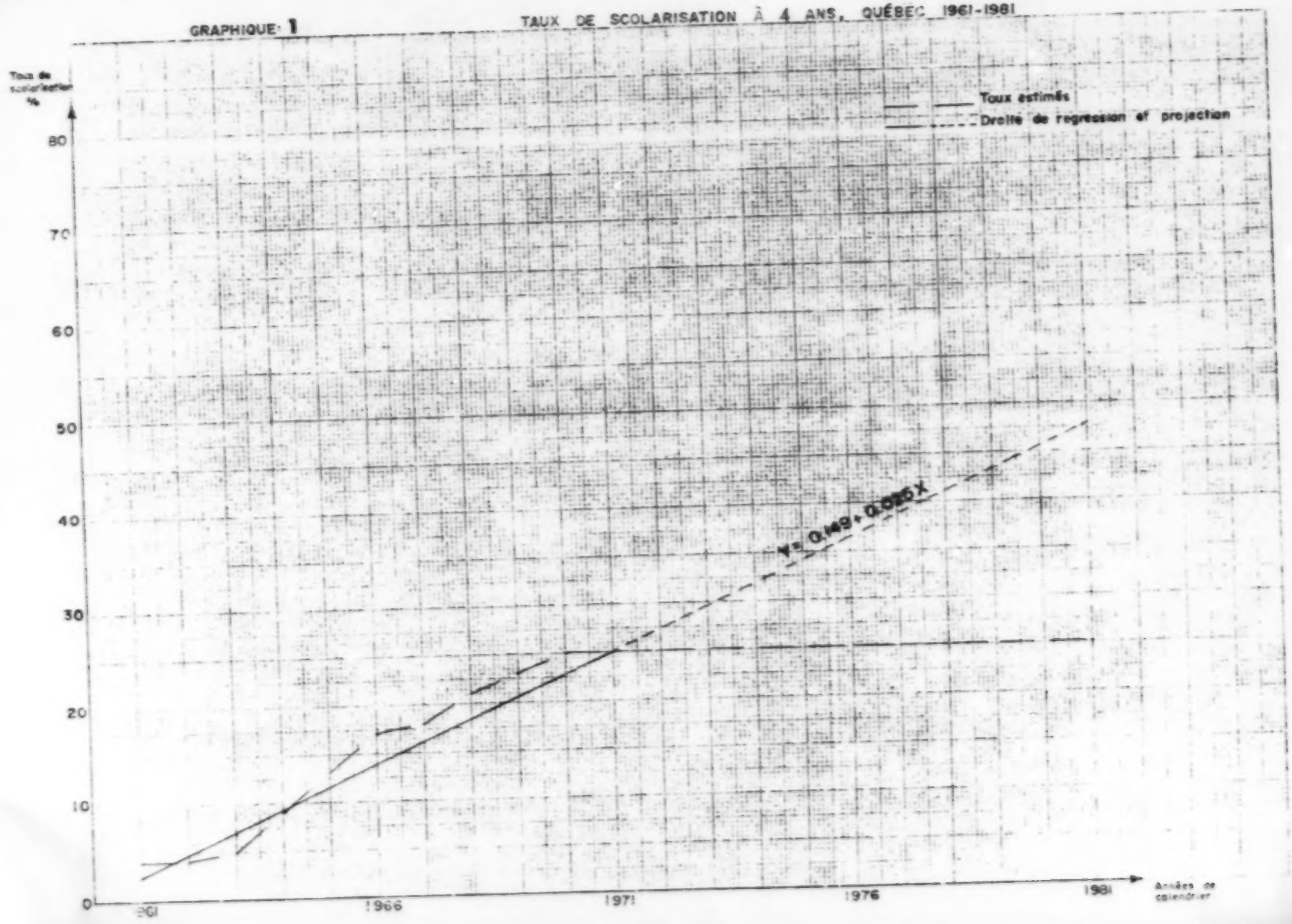
De plus, rien n'indique, même si la tendance était linéaire dans le passé et l'est encore actuellement, qu'elle le sera toujours dans le futur; une telle méthode ne peut prévoir des changements brusques de comportement; bien souvent l'évolution a été si rapide que, si l'on garde constante cette tendance, on dépassera l'unité en quelques années.

Cette méthode pourrait être utilisée pour une prévision à court terme (2 ou 3 ans) et à mesure que de nouvelles données seraient disponibles, il faudrait refaire les calculs pour chaque âge et corriger la prévision si nécessaire.

1- Ibid. page 15.

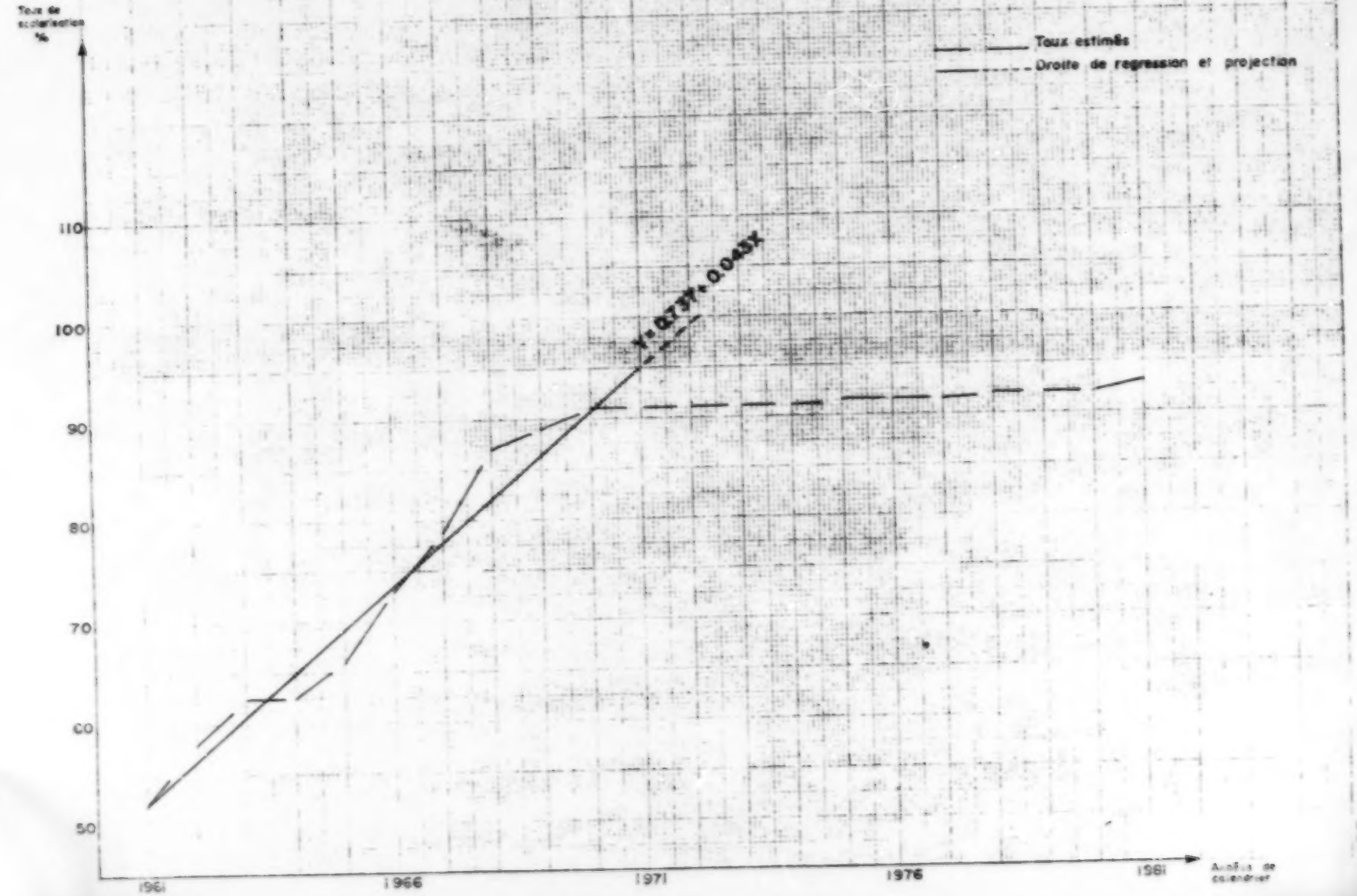
GRAPHIQUE 1

TAUX DE SCOLARISATION À 4 ANS, QUÉBEC 1961-1981



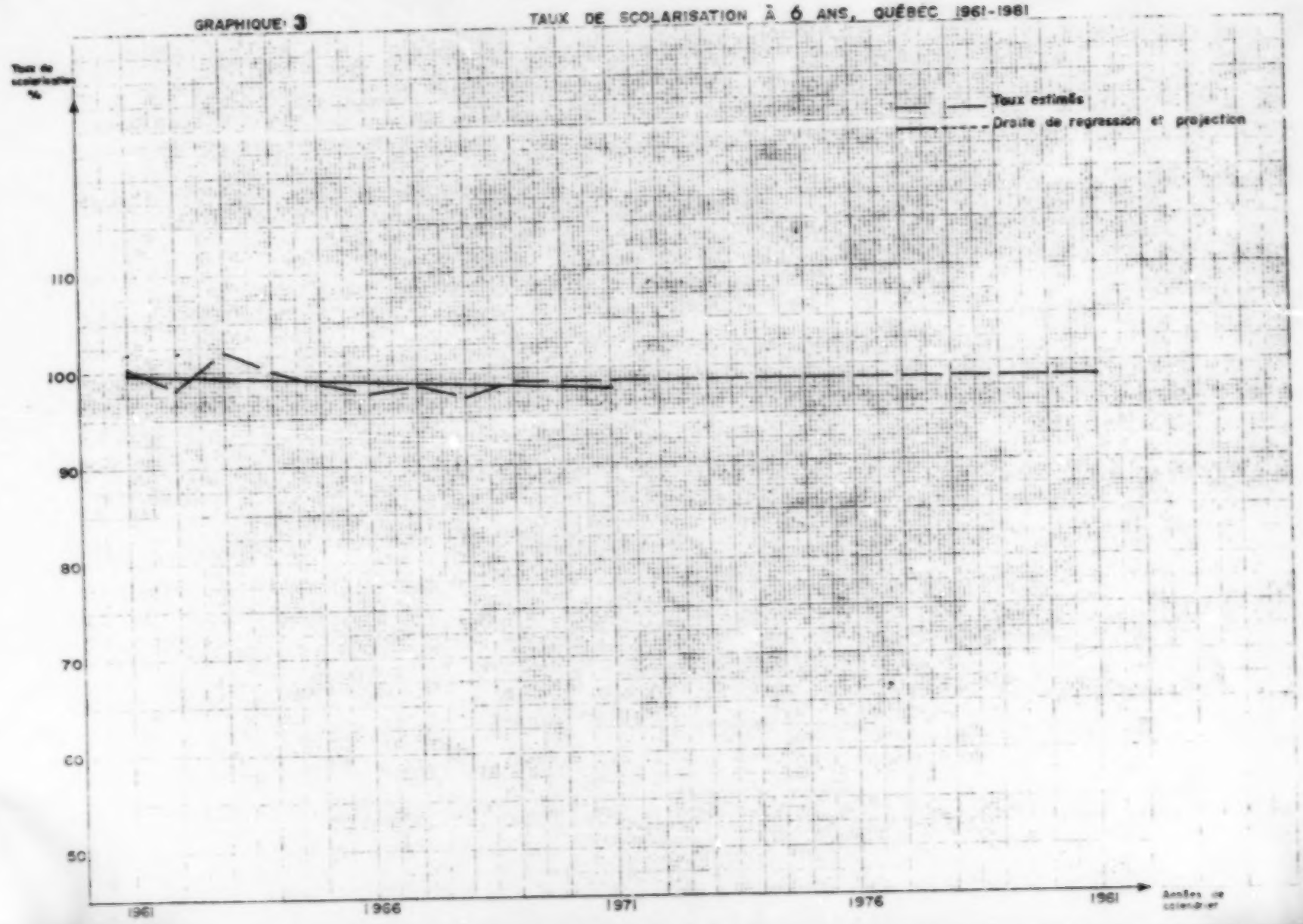
GRAPHIQUE 2

TAUX DE SCOLARISATION À 5 ANS, QUÉBEC 1961-1981



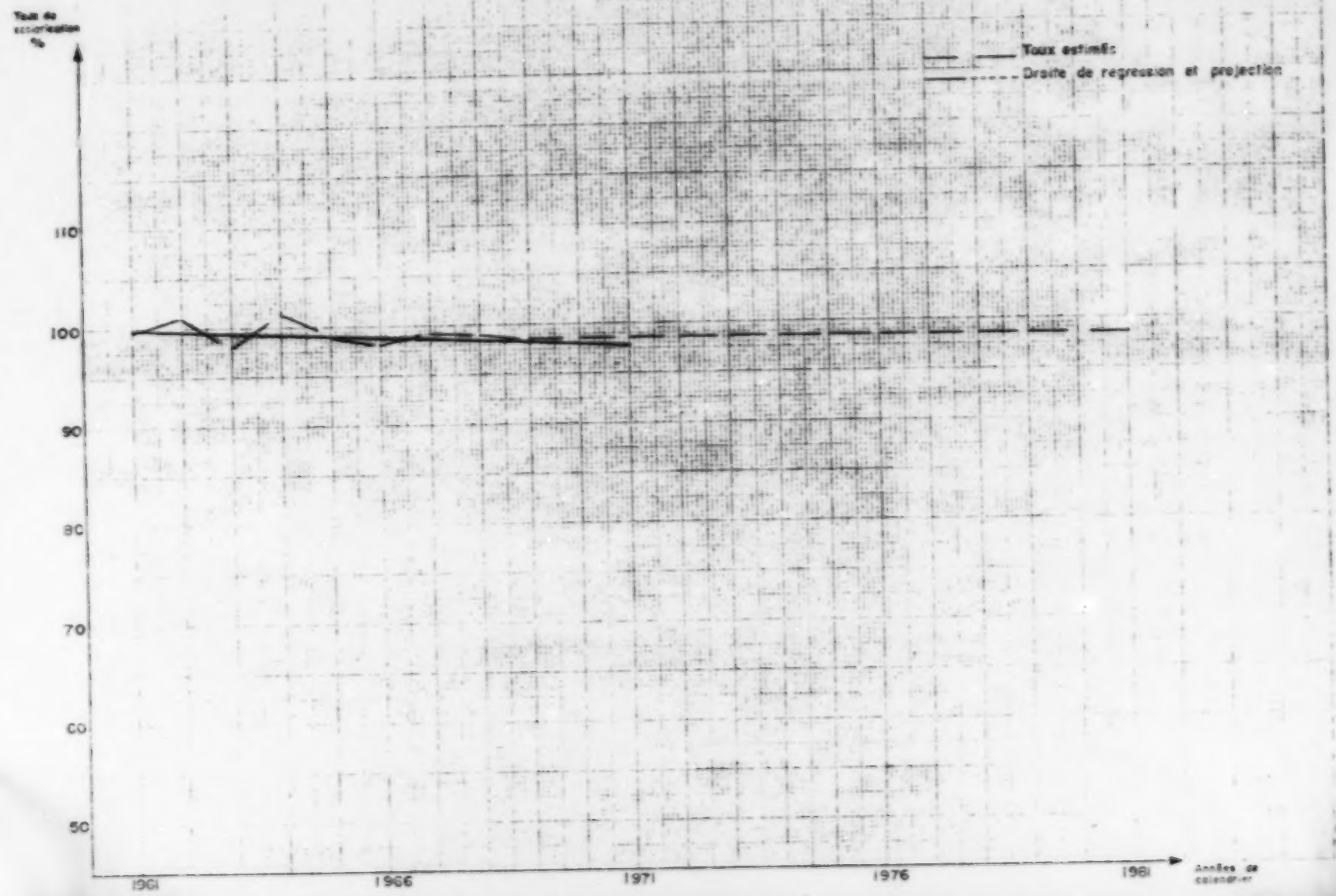
GRAPHIQUE 3

TAUX DE SCOLARISATION À 6 ANS, QUÉBEC 1961-1981



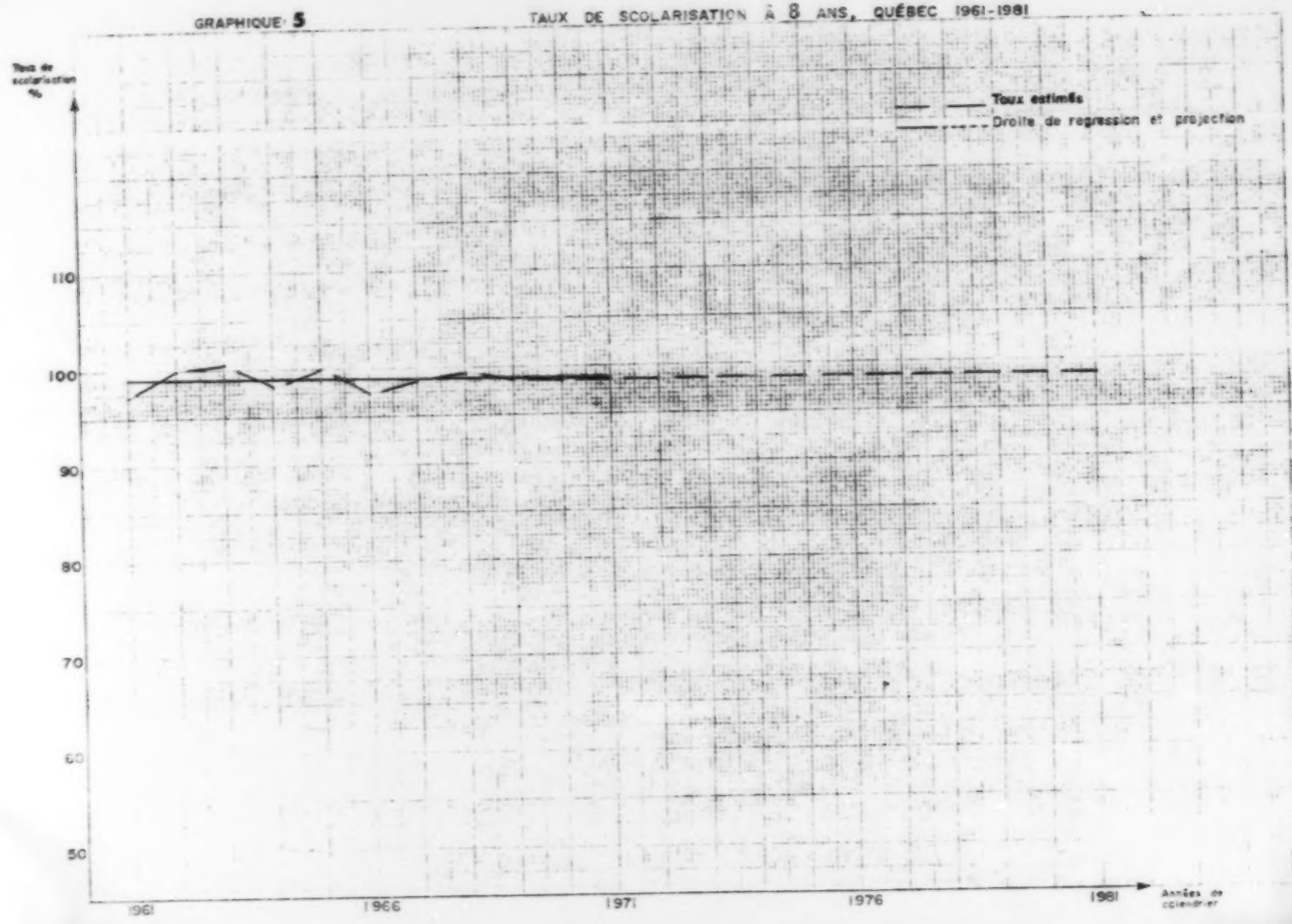
GRAPHIQUE: 4

TAUX DE SCOLARISATION À 7 ANS, QUÉBEC 1961-1981

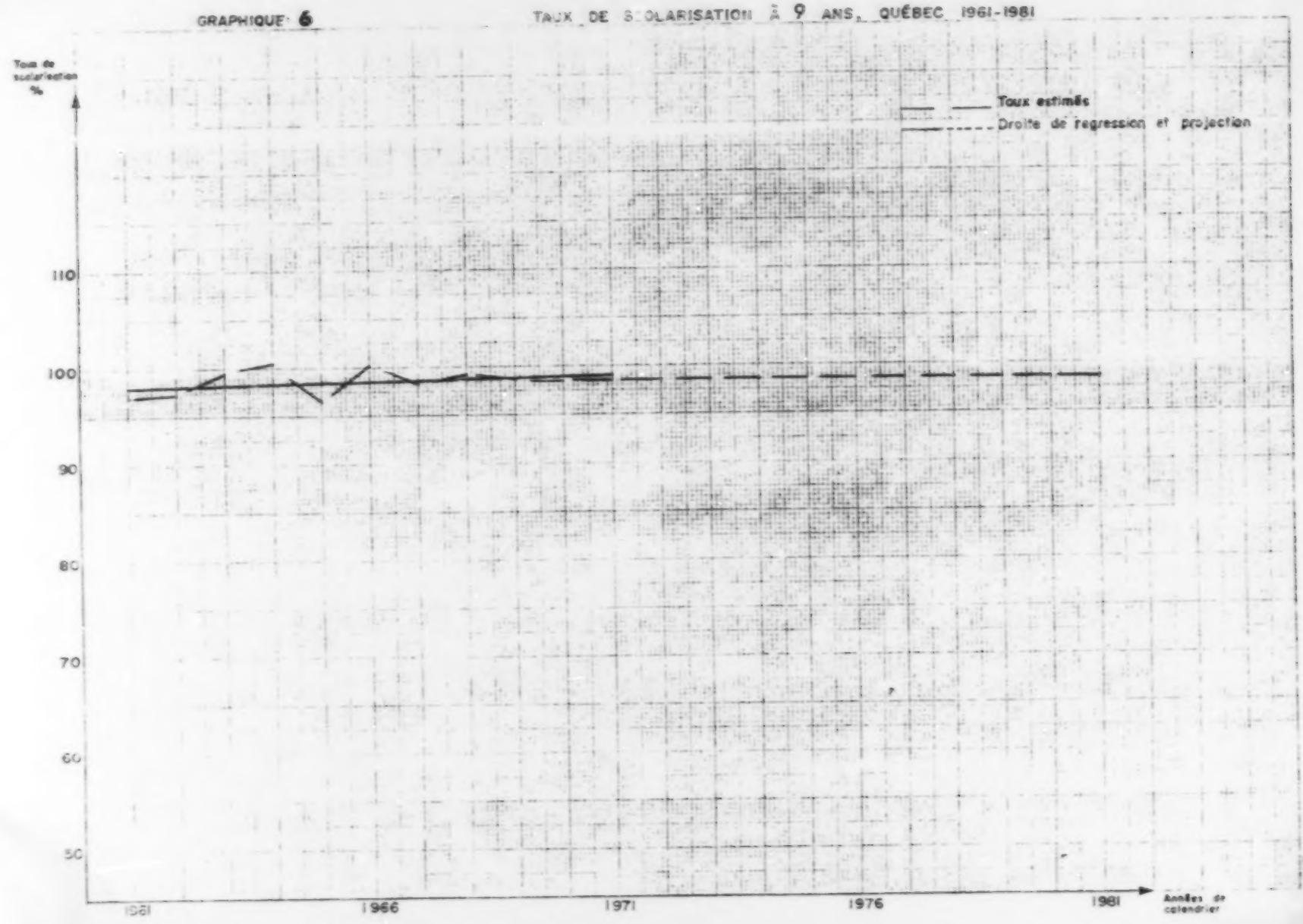


GRAPHIQUE 5

TAUX DE SCOLARISATION À 8 ANS, QUÉBEC 1961-1981

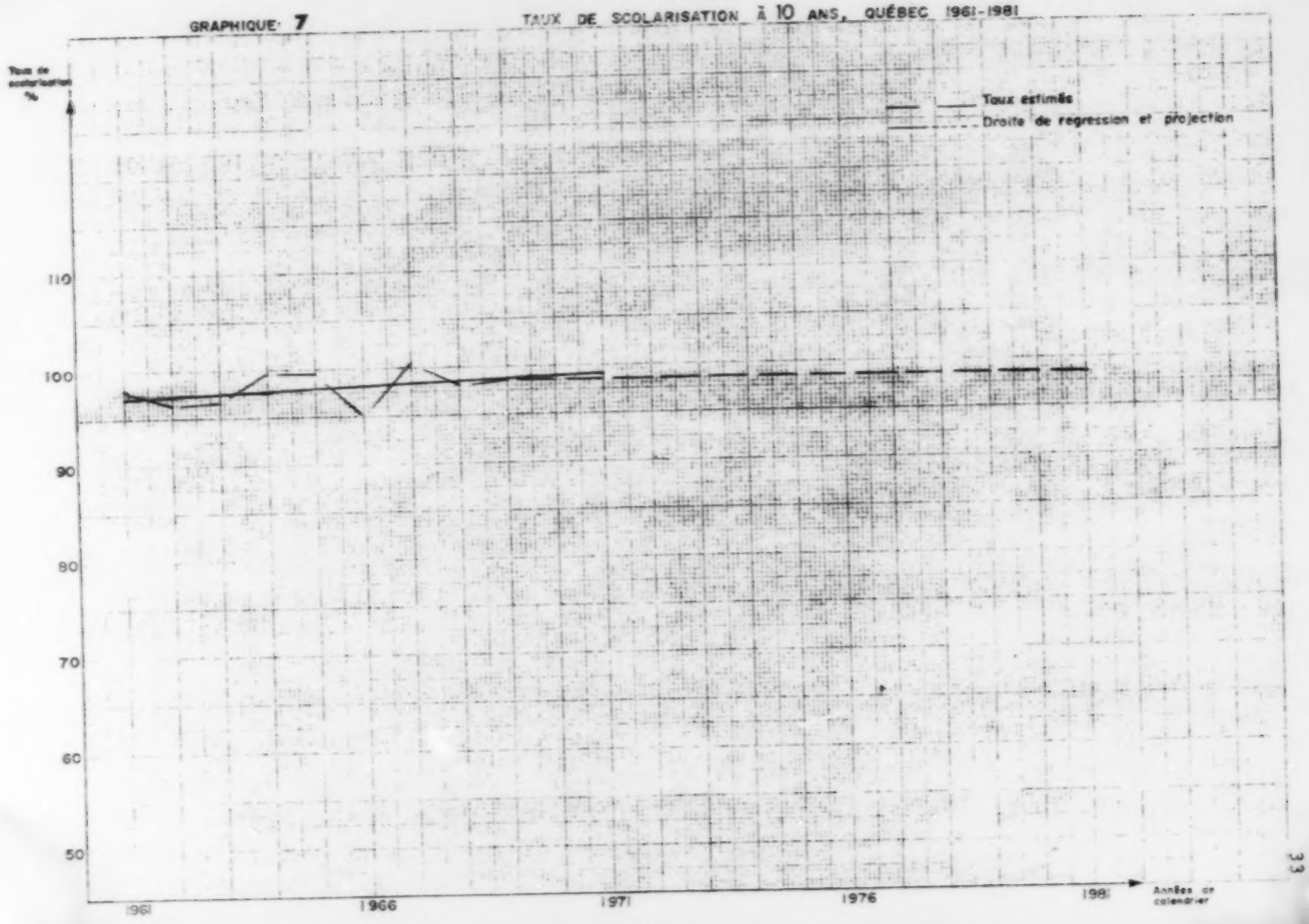


GRAPHIQUE 6 TAUX DE SCOLARISATION À 9 ANS, QUÉBEC 1961-1981



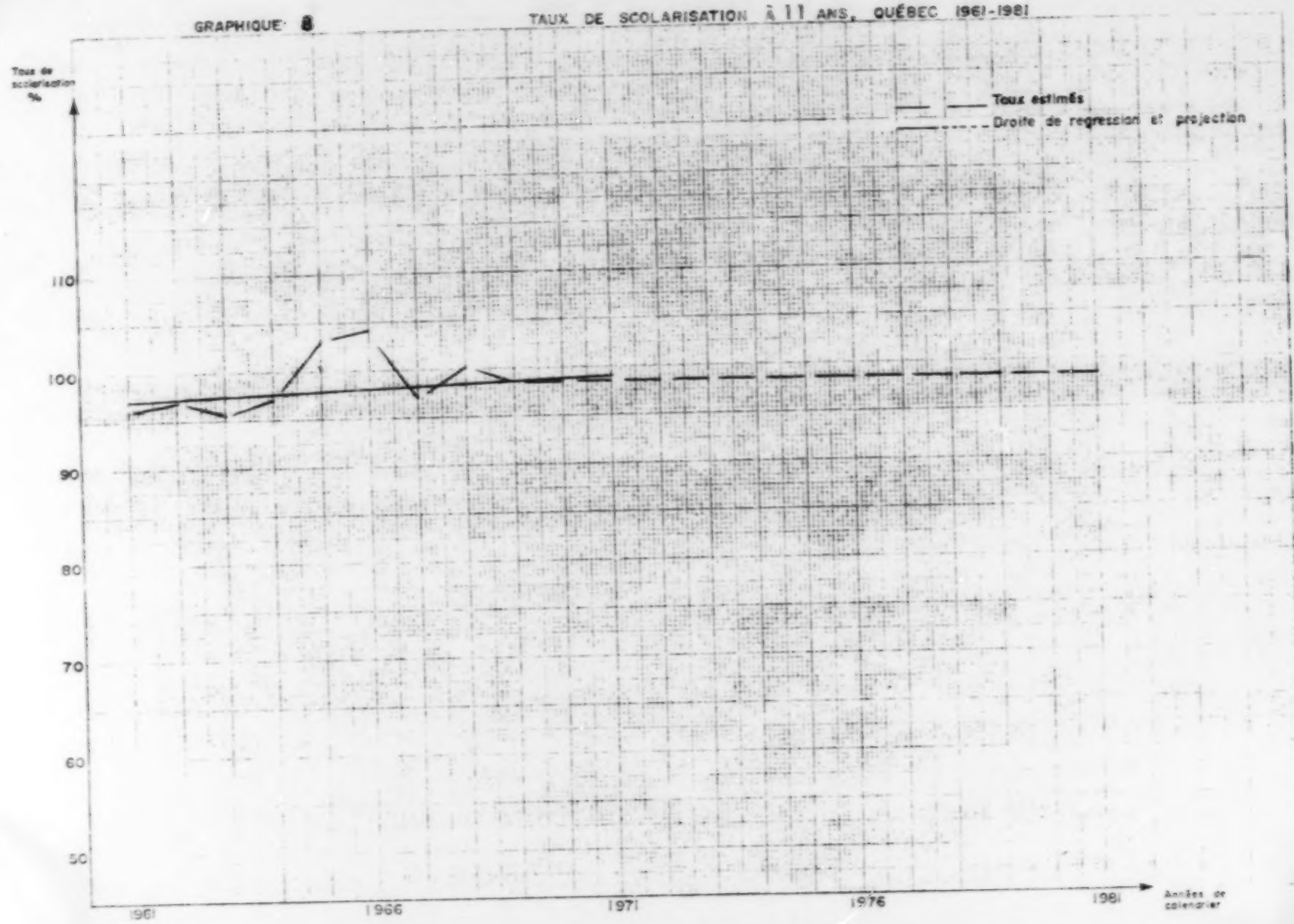
GRAPHIQUE 7

TAUX DE SCOLARISATION À 10 ANS, QUÉBEC 1961-1981



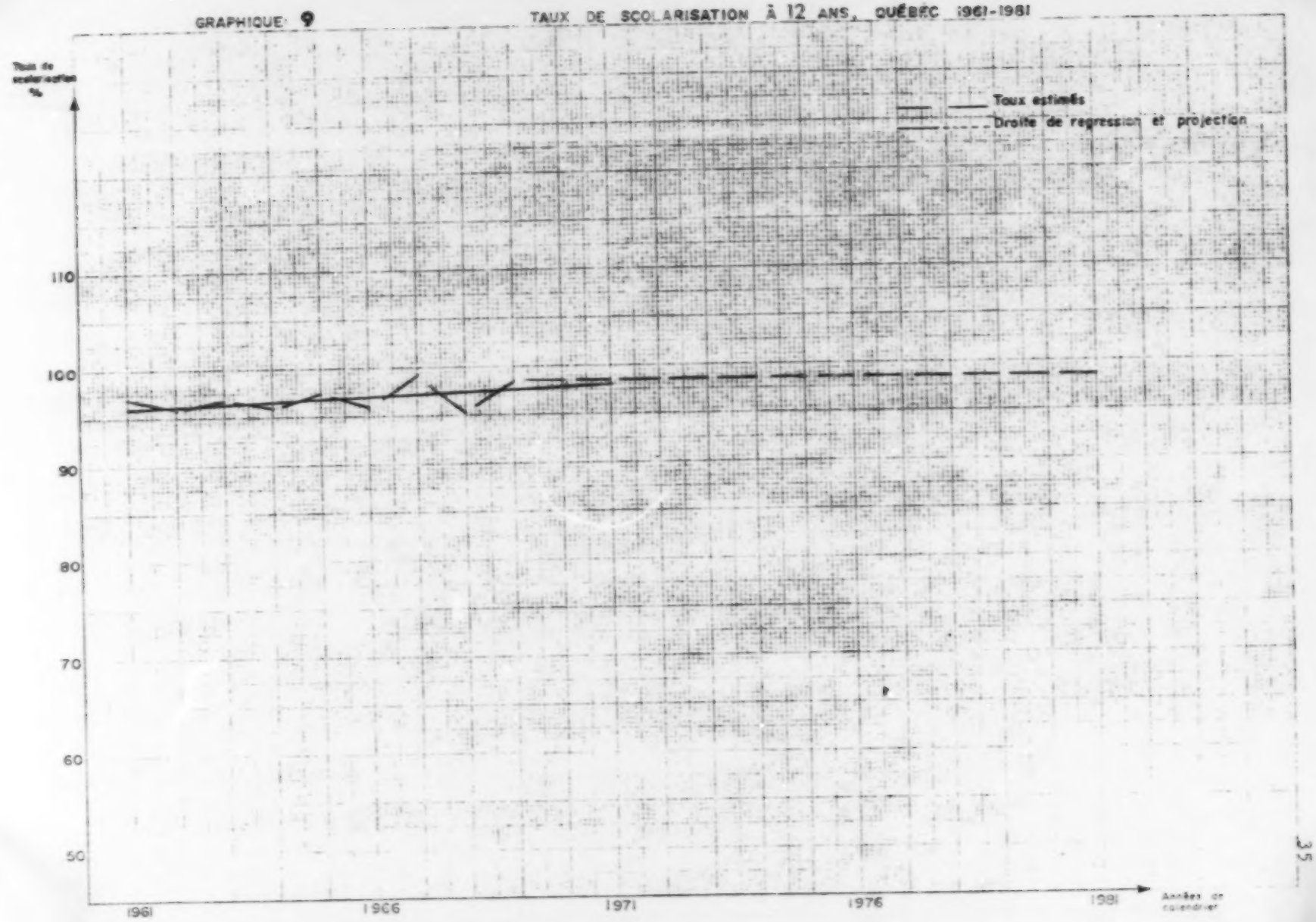
GRAPHIQUE 8

TAUX DE SCOLARISATION À 11 ANS, QUÉBEC 1961-1981



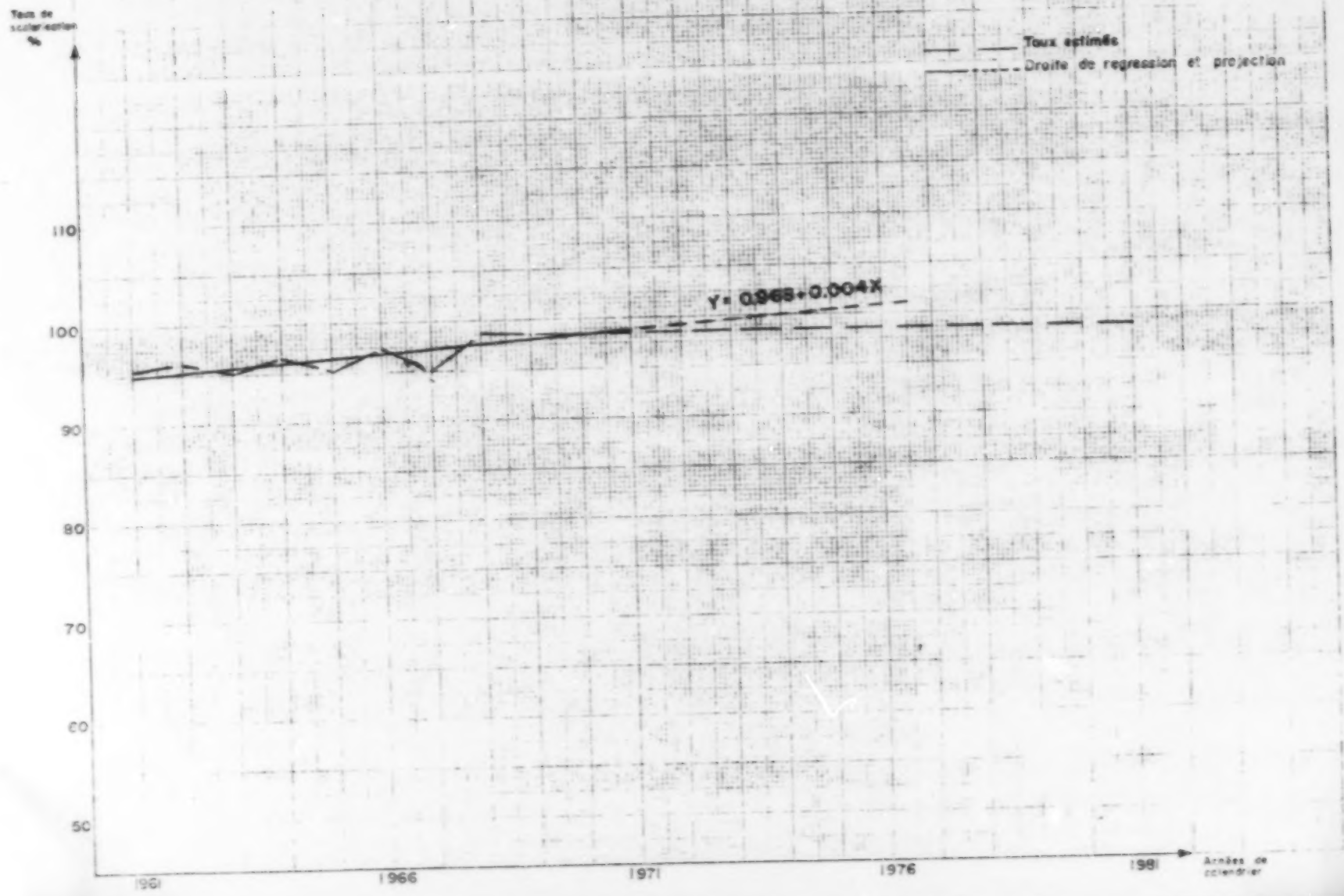
GRAPHIQUE 9

TAUX DE SCOLARISATION À 12 ANS, QUÉBEC 1961-1981



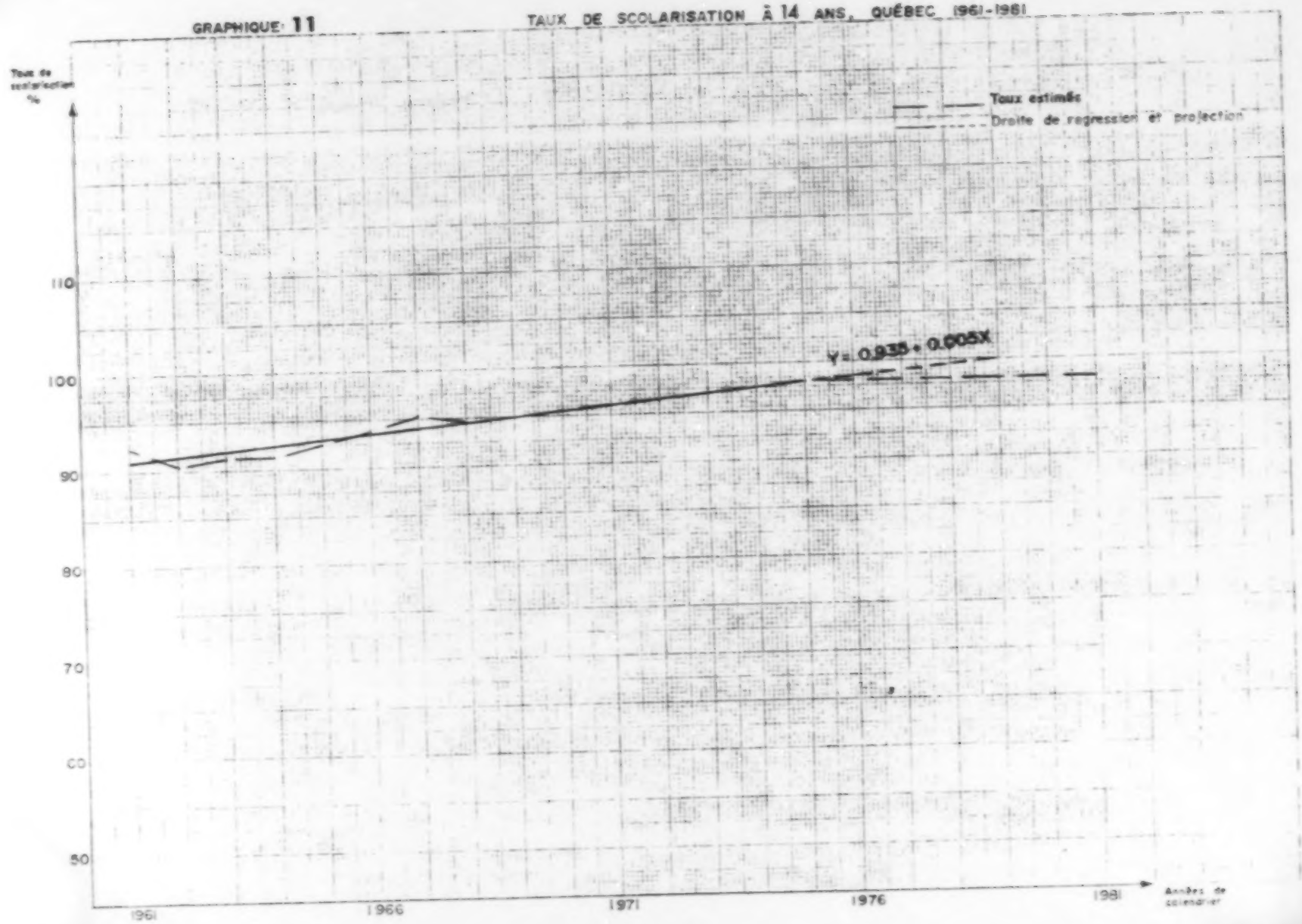
GRAPHIQUE: 10

TAUX DE SCOLARISATION À 13 ANS, QUÉBEC 1961-1981



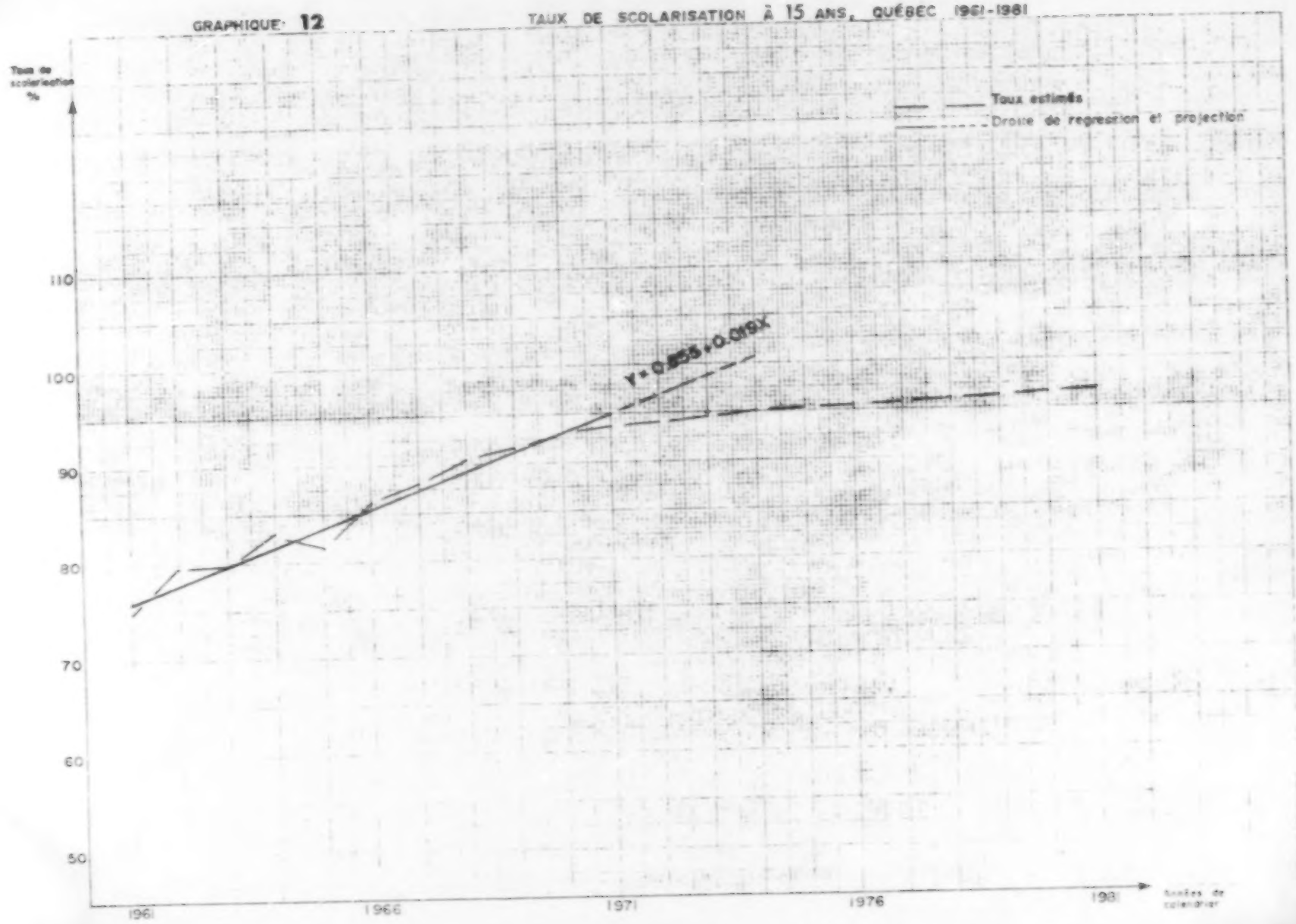
GRAPHIQUE: 11

TAUX DE SCOLARISATION À 14 ANS, QUÉBEC 1961-1981



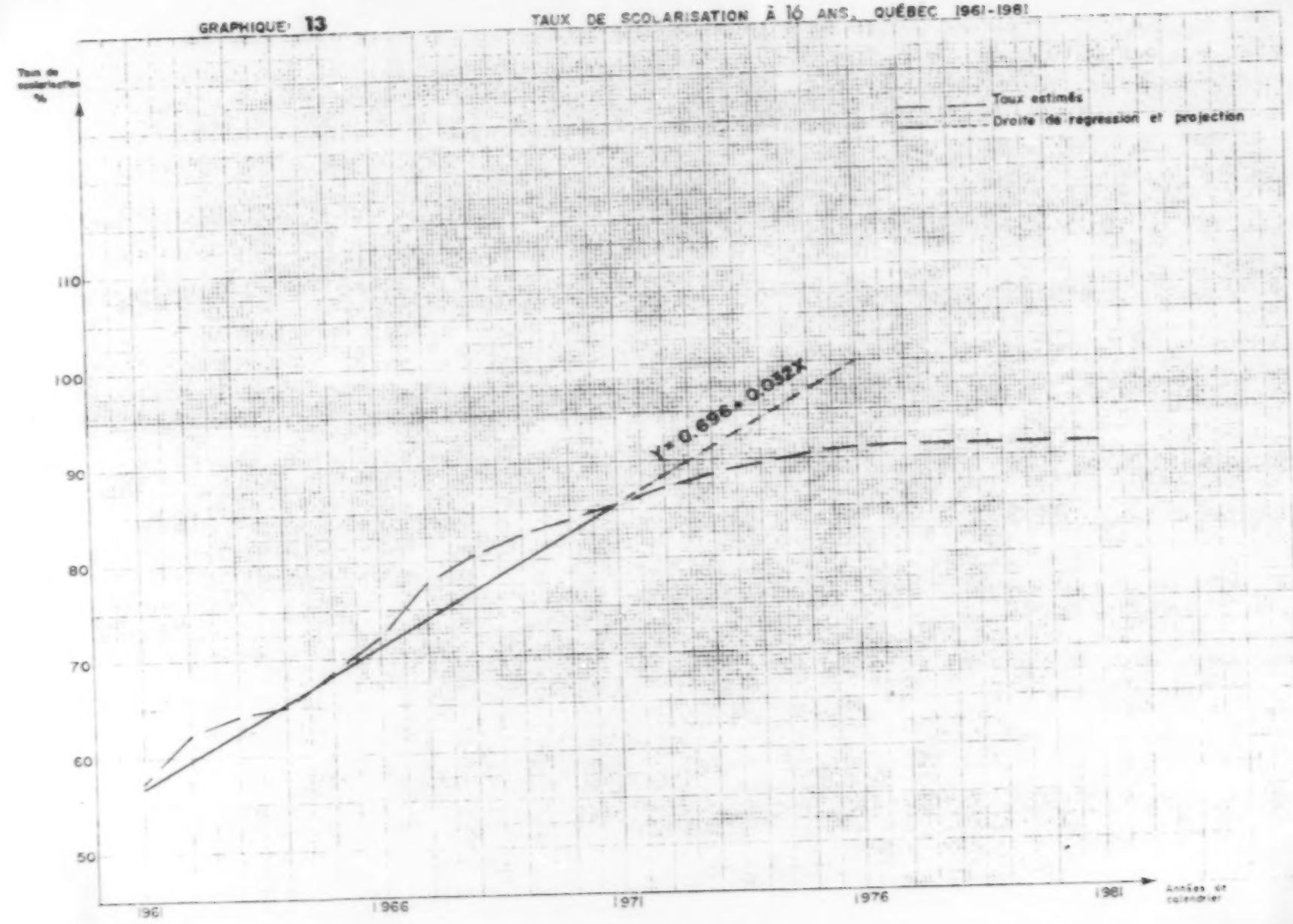
GRAPHIQUE: 12

TAUX DE SCOLARISATION À 15 ANS, QUÉBEC 1961-1981

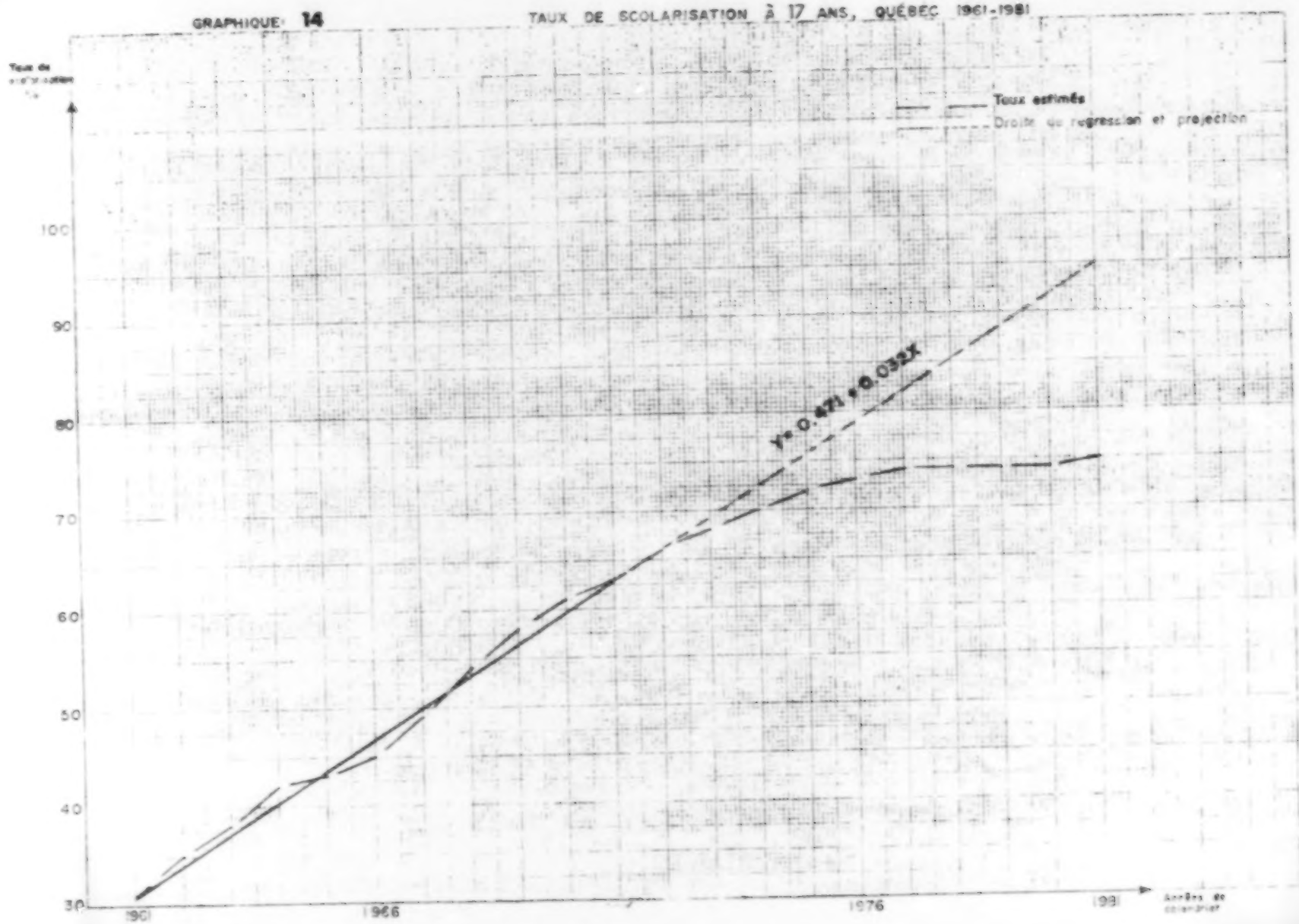


GRAPHIQUE: 13

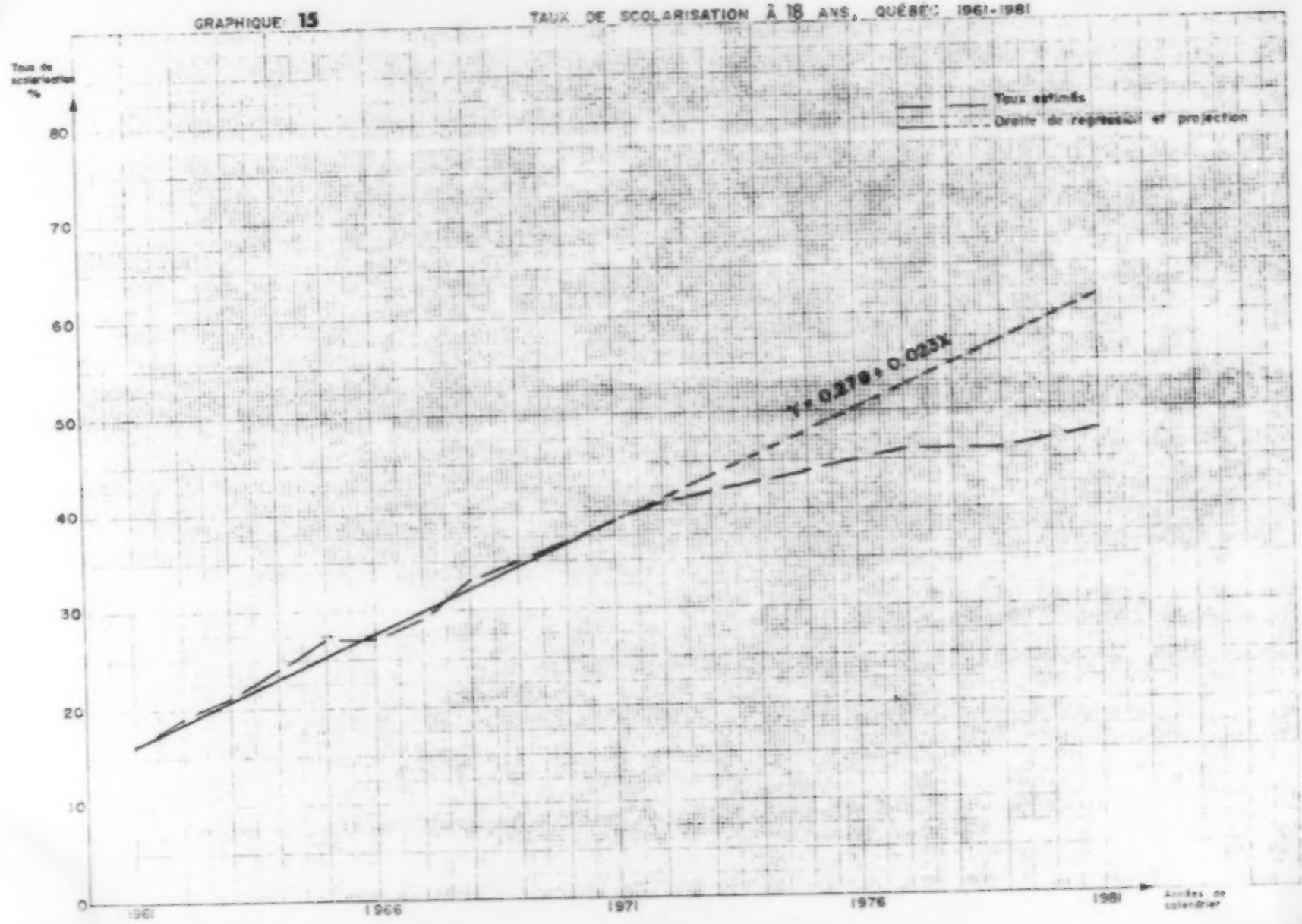
TAUX DE SCOLARISATION À 16 ANS, QUÉBEC 1961-1981



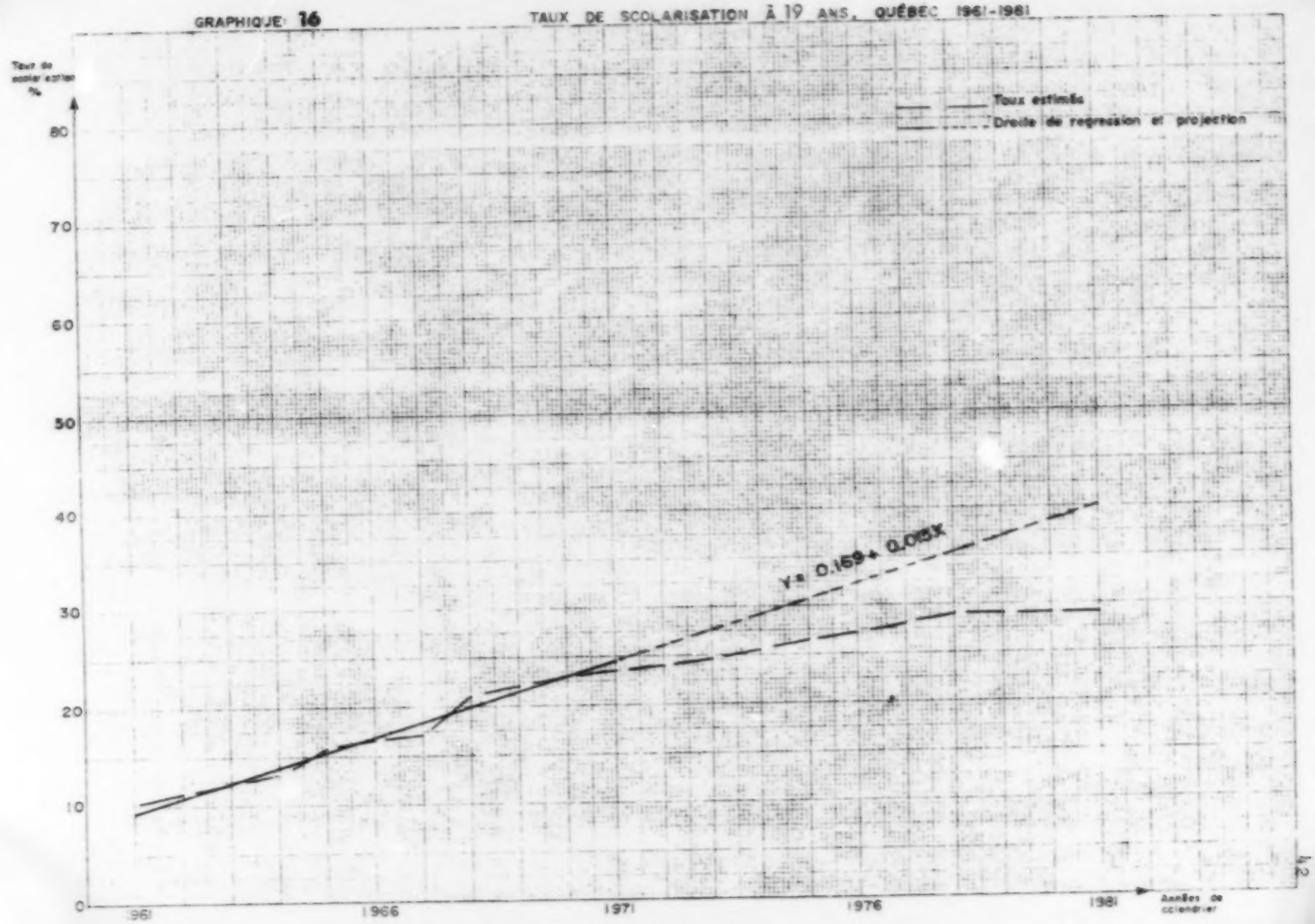
GRAPHIQUE 14 TAUX DE SCOLARISATION À 17 ANS, QUÉBEC 1961-1981



GRAPHIQUE: 15 TAUX DE SCOLARISATION À 18 ANS, QUÉBEC 1961-1981

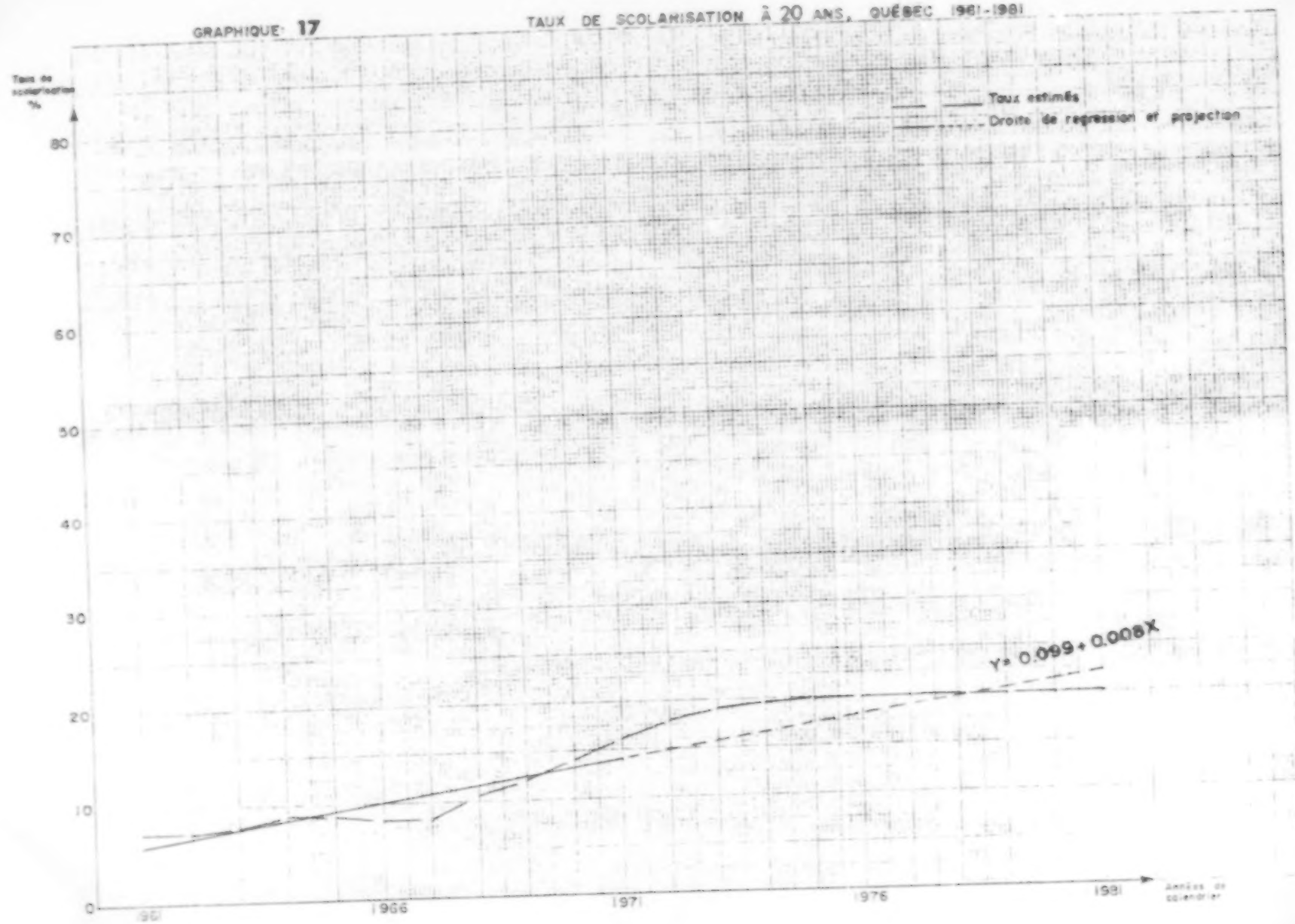


GRAPHIQUE 16 TAUX DE SCOLARISATION À 19 ANS, QUÉBEC 1961-1981

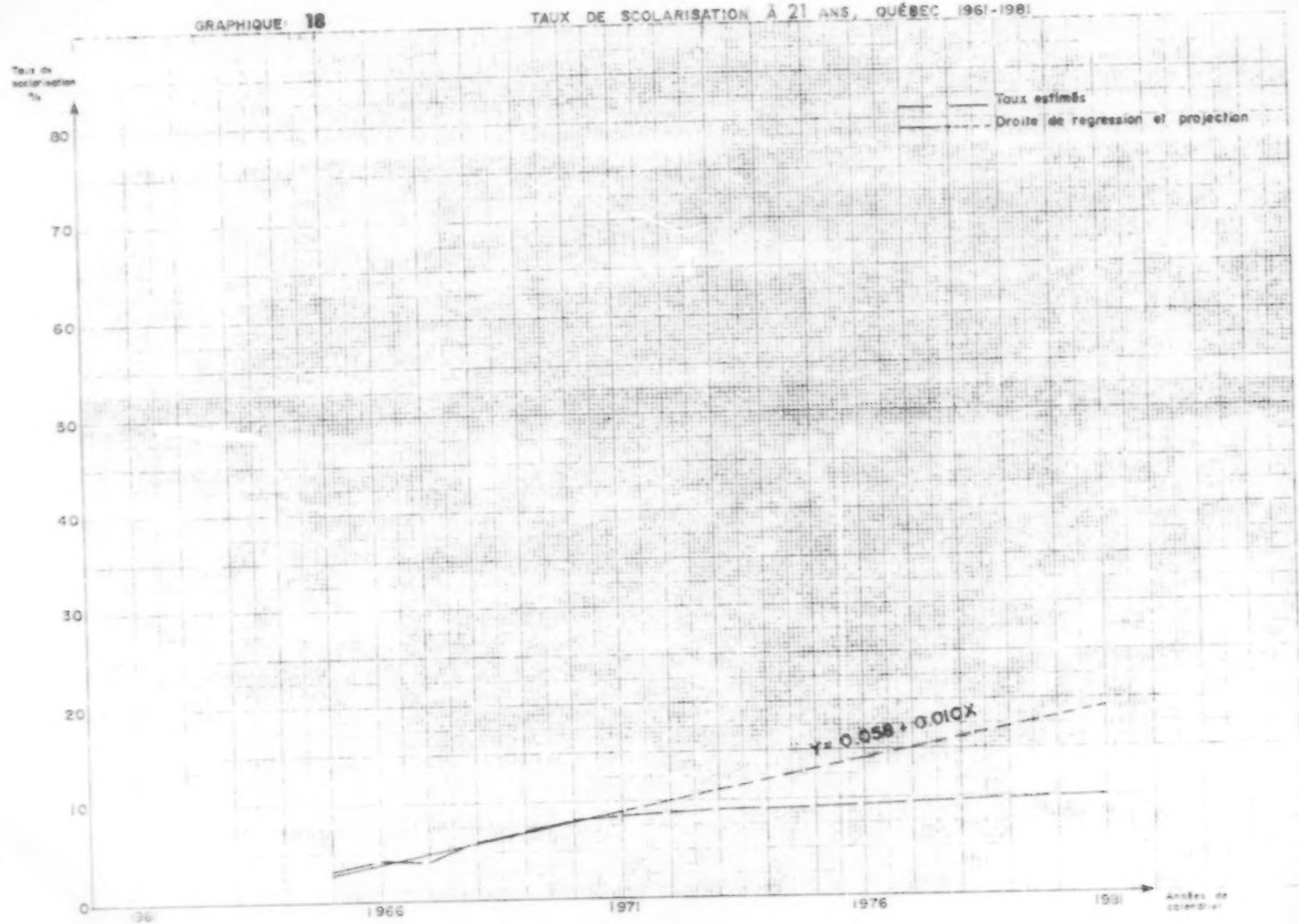


GRAPHIQUE 17

TAUX DE SCOLARISATION À 20 ANS, QUÉBEC 1961-1981

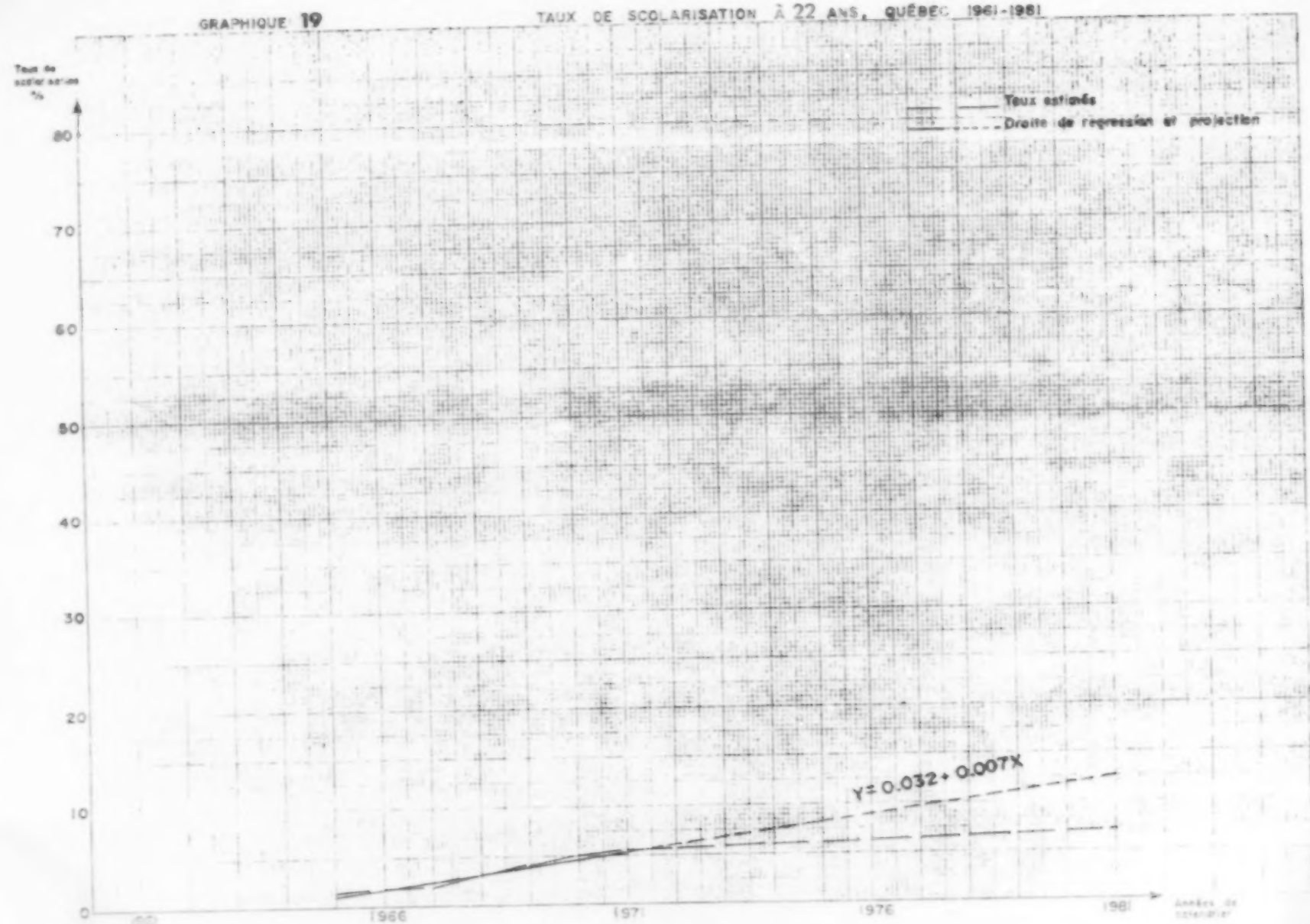


GRAPHIQUE 18 TAUX DE SCOLARISATION À 21 ANS, QUÉBEC 1961-1981



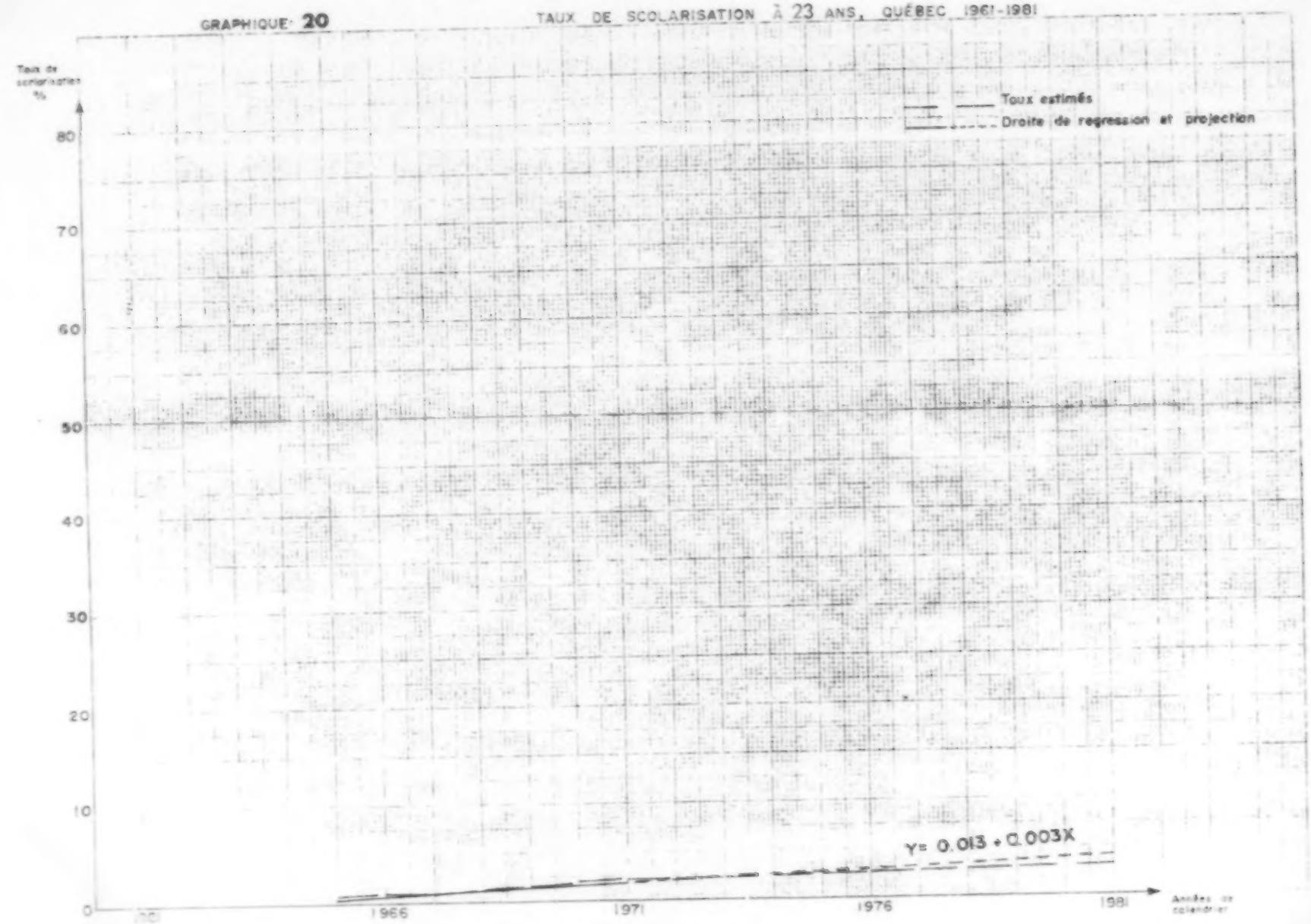
GRAPHIQUE 19

TAUX DE SCOLARISATION À 22 ANS, QUÉBEC 1961-1981



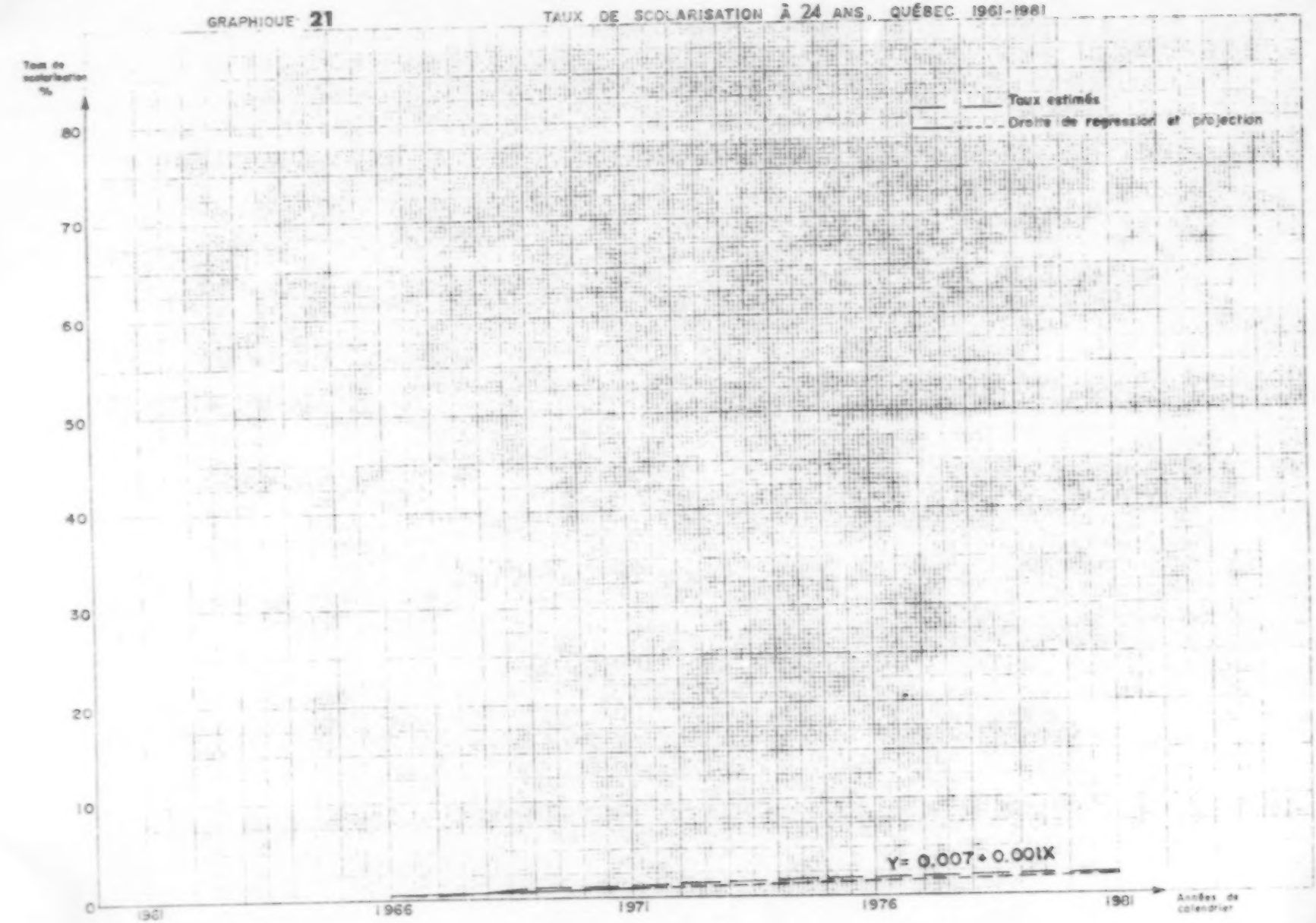
GRAPHIQUE: **20**

TAUX DE SCOLARISATION À 23 ANS, QUÉBEC 1961-1981



GRAPHIQUE 21

TAUX DE SCOLARISATION À 24 ANS, QUÉBEC 1961-1981



SERVICE DE LA REPROGRAPHIE
Septembre 1973